

PENERAPAN HIDDEN MARKOV MODEL (HMM) DAN MEL-FREQUENCY CESPTRAL COEFFICIENTS (MFCC) PADA E-LEARNING BAHASA MADURA UNTUK ANAK USIA DINI

Ubaidi*¹, Nindian Puspa Dewi²

^{1,2}Informatika, Universitas Madura
Email: ¹ubed@unira.ac.id, ²nindianpd@unira.ac.id,
*Penulis Korespondensi

(Naskah masuk: 13 September 2019, diterima untuk diterbitkan: 25 November 2020)

Abstrak

Bahasa Madura adalah Bahasa Daerah yang digunakan di Pulau Madura. Bahasa ini memiliki banyak variasi pengucapan dan dialek. Hal ini menyebabkan Bahasa Madura tidak mudah untuk dipelajari bahkan oleh masyarakat Madura khususnya anak-anak. Saat ini belum ada media pembelajaran Bahasa Madura yang menarik untuk mempelajari Bahasa Madura. Padahal melalui pembelajaran yang menyenangkan diharapkan dapat membantu anak untuk memperoleh kemampuan melatih penerapan pengucapan nama binatang, angka, buah dan benda dalam Bahasa Madura. Karena itulah perlu dibuat *e-learning* Bahasa Madura dengan menerapkan pengenalan pola suara sehingga dapat membantu anak mengenal Bahasa Madura yang memiliki variasi pengucapan untuk objek yang sama. Aplikasi *e-learning* Bahasa Madura untuk anak usia dini digunakan untuk mengenalkan nama objek dalam Bahasa Madura melalui pengenalan pola suara yang diucapkan dibuat dengan menggunakan teknik Mel-Frequency Cepstral Coefficients (MFCC) sebagai ekstrak fitur suara dan Hidden Markov Model (HMM) sebagai teknik pembelajarannya. Penerapan MFCC dan HMM untuk pengenalan pengucapan Bahasa Daerah khususnya Bahasa Madura belum pernah ada sebelumnya, sehingga dengan adanya penelitian ini diharapkan dapat membantu anak usia dini mengenal pengucapan kata Bahasa Madura dengan benar. Pada penelitian ini, sejumlah anak direkam suaranya untuk dijadikan sebagai data training. Data suara yang digunakan adalah data suara yang pengucapan dianggap benar. Skenario percobaan dilakukan dengan menggunakan satu model dan multi model. Setelah dilakukan serangkaian percobaan, hasil penelitian menunjukkan yaitu rata-rata akurasi untuk pengujian sistem dengan satu model yaitu 73% dengan akurasi tertinggi 75% dan rata-rata akurasi untuk pengujian sistem dengan multi model yaitu 80% dengan akurasi tertinggi 81%.

Kata kunci: *e-learning*, pengenalan pola suara, MFCC, HMM

IMPLEMENTATION HIDDEN MARKOV MODEL (HMM) AND MEL-FREQUENCY CESPTRAL COEFFICIENTS (MFCC) IN E-LEARNING OF BAHASA MADURA FOR EARLY CHILDHOOD

Abstract

Bahasa Madura is a regional language used in Madura island. This language has many variations of pronunciation and dialect that makes it not easy to learn, even by the local people especially children. There hasn't been any interesting learning media to learn Bahasa Madura so far. In fact, a fun learning activity is needed to help children to enhance their ability in pronouncing animals' names, numbers, fruits and things in Bahasa Madura. Thus, it's considered important to create Bahasa Madura e-learning by implementing the recognition of voice patterns in order to make it easier for the children to learn Bahasa Madura which has several variations of pronunciation only for one single object. This Bahasa Madura e-learning application for young learners is used to introduce Bahasa Madura vocabularies by recognizing the voice pattern recordings which have been processed through MFCC technique as the extracted voice features and HMM as the learning techniques. The implementation of MFCC and HMM as the learning tool to introduce the pronunciation of regional language vocabularies especially Bahasa Madura has never been done before. Therefore, this research is expected to help the young learners to be able to pronounce Bahasa Madura vocabularies properly. In this study, a number of young learners' voices were recorded and were set as the trial data. Only the proper voice data that were used—voice data that were considered to be pronounced correctly. The trial method was done through one-single model and multi-model. After doing several simultaneous trials, the result showed the accuracy level. The average

accuracy level for one-single model system was 73% (with the highest accuracy reached 75%) and the average accuracy level for multi-model system was 80% (with the highest accuracy reached 81%).

Keywords: *e-learning*, voice pattern recognition, MFCC, HMM.

1. PENDAHULUAN

E-learning merupakan Media pembelajaran yang memanfaatkan teknologi sistem informasi dalam penerapannya. Selain dapat digunakan secara mandiri, penggunaan *e-learning* yang memanfaatkan penggunaan audio dan visual dalam pembelajarannya dapat mengurangi rasa bosan dalam proses belajar mengajar. Bentuk pembelajaran yang menyenangkan diharapkan dapat memudahkan, memotivasi dan meningkatkan semangat penggunanya untuk belajar.

Banyak *e-learning* yang telah dibuat namun belum ada yang menerapkan pengenalan pola suara untuk proses evaluasi pembelajaran. Pengenalan pola dapat dibuat dengan menggunakan MFCC dan HMM sebagai teknik pengenalan pola suaranya. Pengenalan suara merupakan salah satu bagian dari bidang aplikasi yang memungkinkan suatu perangkat untuk mengenali dan memahami kata-kata yang diucapkan dengan mengimplementasikan sebuah algoritma ke dalam sistem (Abdullah, Chow & Sin, 2003). Pengenalan suara dilakukan dengan melakukan ekstraksi sinyal suara-suara (Sanjaya, 2017) dan pencocokan fitur suara yang masuk ke dalam sistem. Salah satu teknik ekstraksi suara adalah MFCC yang merupakan salah satu metode terbaik dan sangat umum digunakan untuk ekstraksi fitur sinyal suara (Mehta & Dabhade, 2013). Sedangkan untuk pengenalan suara salah satunya adalah Hidden Markov Model (Permana, 2011). Metode Hidden Markov (HMM) adalah pendekatan yang paling cocok untuk pengenalan suara karena efisien, kuat, dan mengurangi waktu dan kompleksitas (Naziya & Deshmukh, 2015).

Banyak penelitian tentang *e-learning* yang telah dilakukan seperti Implementasi Game Based Learning pada Pembelajaran Bahasa Inggris (Dewi & Listiowarni, 2019), E-learning English For Children Aplikasi Mobile Mengenal Alat Transportasi Berbasis Android (Yundatama & Parno, 2015) yaitu pembuatan *e-learning* sebagai media pembelajaran Bahasa Inggris dengan menggunakan android dan Rancang Bangun Aplikasi E-learning untuk Pembelajaran Agama Islam berbasis Android (Cantika, 2017) yang merupakan pemanfaatan *e-learning* untuk pembelajaran Agama Islam. Sedangkan untuk penelitian yang membahas tentang pengenalan pola suara yaitu Automatic Speech And Speaker Recognition By MFCC, HMM And Vector Quantization (Deshmukh & Bachute, 2013), Feature Extraction Using MFCC (Gupta dkk, 2013), Speech-To-Text Conversion (STT) System Using Hidden Markov Model (HMM) (Mon & Tun, 2015) dan Voice Recognition using HMM with MFCC for Secure ATM (Iqbal, Mahboob & Khiyal, 2011),

MFCC and Its Applications in Speaker Recognition (Tiwari, 2010).

Pada penelitian ini *e-learning* akan digunakan sebagai media pengenalan dalam Bahasa Madura untuk anak usia dini dengan menggunakan MFCC dan HMM. Pemilihan Bahasa Madura sebagai objek pembelajaran karena semakin berkurangnya penggunaan Bahasa Madura terutama dikalangan generasi muda di Madura. Menurut Peneliti dari LIPI, Prof Dr Mien A Rifai, menyatakan di antara bahasa-bahasa daerah yang ada di Indonesia, Bahasa Madura merupakan salah satu bahasa terbesar keempat setelah Bahasa Jawa, Bahasa Indonesia dan Bahasa Sunda dari segi jumlah penuturnya. Tapi perkembangannya dalam dasawarsa terakhir ini sangat memprihatinkan. Berkurangnya penggunaan Bahasa Madura salah satunya akibat adanya kebijakan dari pemerintah pusat yang mengharuskan Bahasa Indonesia sebagai satu-satunya bahasa pengantar dalam bidang pendidikan. Karena itulah pengenalan Bahasa Madura harus dilakukan sejak usia dini sehingga tetap dapat dikenal oleh generasi muda Madura.

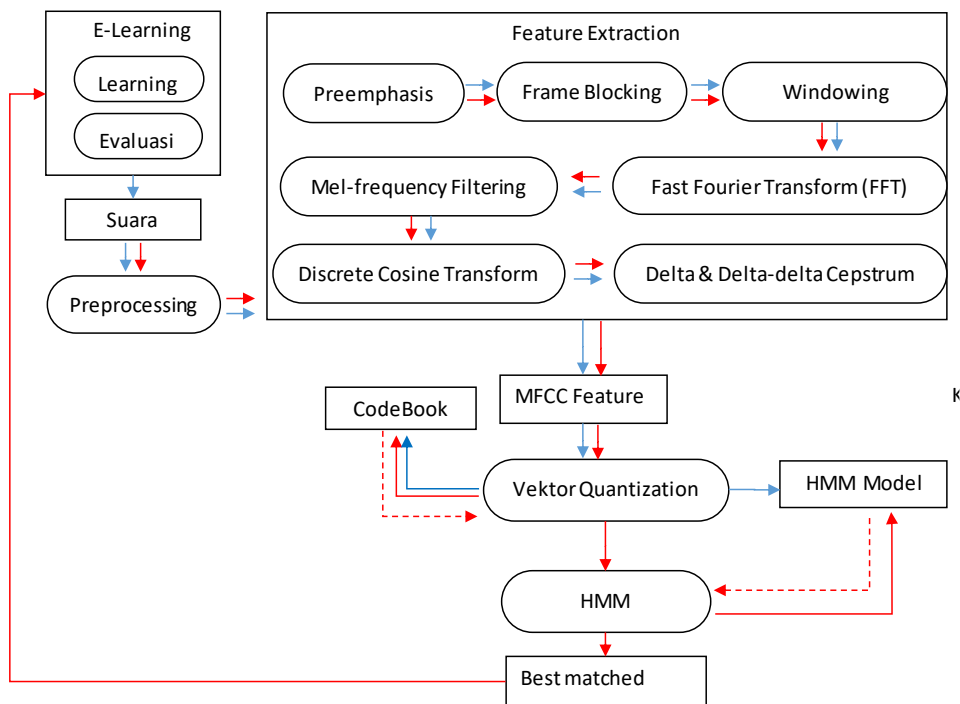
Dari sudut pandang linguistik murni, Bahasa Madura memiliki 4 dialek yaitu (1) dialek Sumenep, (2) dialek Pamekasan, (3) dialek Bangkalan, dan (4) dialek Kangean. Masing-masing dialek digunakan di daerah yang berbeda di Madura (Sofyan, 2007). Perbedaan menonjol dari diantara masing-masing dialek itu adalah dalam hal pengucapan, terutama yang berkenaan dengan prosodi dan intonasi. Selain itu dalam Bahasa Madura juga terdapat enam vokal yaitu /a/, /i/, /u/, /e/, /ə/, dan /ɔ/ (Sofyan, 2017) yang membuat pengucapan Bahasa Madura tidak semudah Bahasa Indonesia, sehingga banyak orang Madura yang juga masih belum bisa membaca tulisan dengan kosa kata Bahasa Madura.

Pembelajaran Bahasa Madura dengan menerapkan *e-learning* belum pernah ada sebelumnya. Padahal metode pembelajaran ini dapat menarik perhatian anak mengingat sulitnya belajar Bahasa Madura khususnya untuk membantu dalam pengenalan pengucapan kata Bahasa Madura. Karena itulah perlu dicoba untuk membuat *e-learning* dengan menerapkan pengenalan pola suara untuk pembelajaran Bahasa Madura. Penelitian ini diharapkan dapat membantu anak belajar cara pengucapan nama suatu objek dalam Bahasa Madura yang memiliki variasi pengucapan dan dialek. Dalam penelitian ini dibatasi pada pengenalan pola suara dengan menggunakan dialek Pamekasan serta menggunakan suara anak yang cukup baik pengucapannya.

Pada bagian selanjutnya akan dibahas bagaimana penelitian ini dilakukan dimulai dengan *preprocessing*, langkah-langkah ekstraksi fitur suara dengan MFCC sampai pengenalan pola suara dengan HMM. Untuk uji coba dilakukan dengan menggunakan dua skenario uji coba yaitu menggunakan satu model dan multi model. Uji coba dengan akurasi tertinggi akan digunakan untuk penelitian selanjutnya.

2. METODE PENELITIAN

Secara umum sistem yang dibuat dalam penelitian ini dibagi menjadi empat bagian yaitu *e-learning*, *preprocessing*, ekstraksi fitur, pelatihan, dan pengenalan pola suara. Berikut blok diagram penelitian dapat dilihat pada gambar 1 yang masing-masing bagian akan dijelaskan lebih lanjut dalam sub bab selanjutnya.



Gambar 1. Blok Diagram Penelitian

2.1. Preprocessing

Suara yang masuk ke dalam sistem akan mengalami proses *preprocessing* yang merupakan proses untuk membuang *unvoiced*. Dalam pengembangan sistem pengenalan suara, *preprocessing* merupakan fase pertama dari fase-fase lain dalam pengenalan suara. *Preprocessing* menyesuaikan atau memodifikasi sinyal suara, $x(n)$, sehingga akan lebih dapat diterima untuk kemudian dilakukan ekstraksi fitur. *Preprocessing* adalah pemrosesan untuk memeriksa sinyal suara, $x(n)$ karena bisa saja sinyal suara dipengaruhi lingkungan sekitar, $d(n)$, yang merupakan gangguan aditif (Ibrahim, Odiketa & Ibiyemi, 2017).

2.1.1. Konversi Analog menjadi Digital

Konversi analog menjadi digital merupakan proses untuk mengubah sinyal analog menjadi sinyal digital (amplitudo). Proses diperlukan sehingga suara dapat diolah lebih lanjut oleh program. Adapun format sinyal yang digunakan dalam penelitian ini yaitu 22050 Hz, 16-bits dan Mono Channel.

2.1.2. Endpoint Detection dan Silence Removal

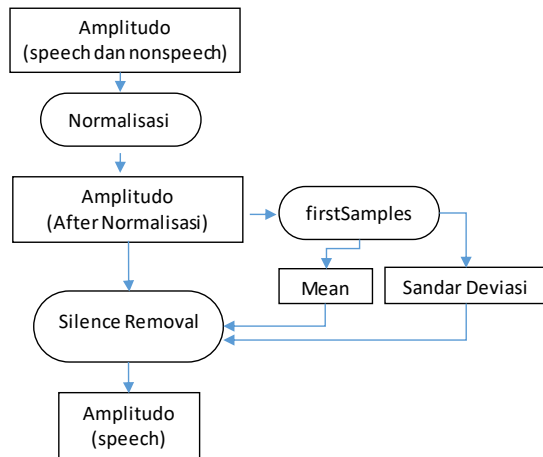
Preprocessing yang digunakan pada penelitian ini adalah dengan *silence removal* dan *endpoint detection* (Saha, Chakroborty, & Senapati 2014). Biasanya 200 ms pertama atau lebih (kami menggunakan 32768 sampel untuk laju pengambilan sampel 22050 samples/detik) dari rekaman suara terdiri dari sinyal yang tidak ada suaranya (*unvoiced*) atau suara dari lingkungan. Hal ini karena biasanya *speaker* membutuhkan beberapa waktu untuk membaca atau berpikir ketika perekaman dimulai. Pada Gambar 2 berikut menunjukkan proses *Endpoint Detection* dan *Silence Removal*.

Dalam *preprocessing*, sinyal suara yang sudah berupa amplitudo akan dinormalisasi terlebih dahulu. Dalam penelitian ini akan diambil 200 ms bagian suara yang kemudian dicari mean dan standar deviasinya.

Persamaan (1) digunakan untuk menghitung mean dan persamaan (2) untuk menghitung standar deviasi.

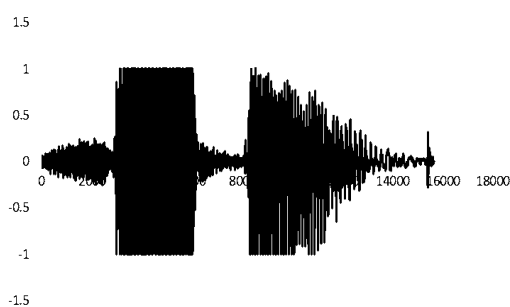
$$\text{mean}(\mu) = \frac{\sum \text{Normalisasi}}{n} \quad (1)$$

$$\text{standar deviasi}(\delta) = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n}} \quad (2)$$



Gambar 2. Blok Diagram Endpoint Detection dan Silence Removal

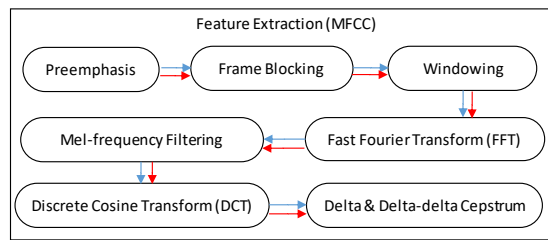
Pada persamaan (1) dan (2), x_i adalah Nilai amplitudo yang sudah dinormalisasi dan n adalah jumlah sampel yang digunakan. Dikatakan *voiced* jika $(|s[n] - \mu|)/\delta > 2$ dan *unvoiced* jika $(|s[n] - \mu|)/\delta < 2$. Dengan $s[n]$ adalah sampel amplitudo ke- n . Pada Gambar 3 tampak perbedaan grafik amplitudo sesudah dan sebelum dilakukan *Silence Removal*. Hal ini karena amplitudo yang tidak memiliki suara (*unvoiced*) akan dihilangkan dan sebaliknya yang memiliki suara (*voiced*) dipilih sebagai data untuk proses selanjutnya.



Gambar 3. Amplitudo sesudah dilakukan Silence Removal

2.2. Feature Extraction dengan MFCC

Hasil dari *preprocessing* ini akan diekstraksi dengan MFCC untuk menghasilkan MFCC *feature* (Panchal, 2010). MFCC adalah metode yang paling banyak digunakan dalam pemrosesan suara, karena dianggap cukup baik dalam merepresentasikan sinyal suara (Sanjaya & Salleh, 2014). Pada Gambar 4 menunjukkan Blok diagram proses *Feature Extraction* dengan MFCC.

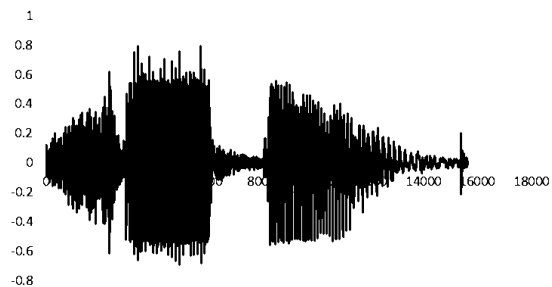


Gambar 4. Blok Diagram Feature Extraction dengan MFCC

Proses pertama dari *feature extraction* adalah *pre-emphasis* yang berfungsi untuk meningkatkan frekuensi. Fungsi dari *pre-emphasis* hampir sama dengan *high-pass filter* yaitu untuk melewatkan komponen frekuensi tinggi dari sinyal suara. Default nilai α yang digunakan dalam penelitian ini adalah 0.95 dengan melakukan perhitungan pada persamaan (3).

$$y[n] = s[n] - \alpha s[n - 1] \quad (3)$$

pada persamaan(3), $y[n]$ adalah sinyal hasil *pre-emphasis filter*, $s[n]$ adalah sinyal sebelum *pre-emphasis filter* dan α merupakan konstanta *pre-emphasis filter*. Pada umumnya nilai α yang paling sering digunakan adalah antara 0.9 sampai 1.0. Gambar 5 menunjukkan contoh hasil *pre-emphasis*.



Gambar 5. Amplitudo hasil Preemphasis

Frame blocking akan memecah sinyal suara menjadi beberapa *frame* dimana masing-masing *frame* kemudian dikenai *windowing* untuk menyatukan setiap bagian dalam *frame* agar tidak terpecah. Perhitungan banyak *frame* yang digunakan persamaan(4) berikut.

$$\text{number Of Frames} = 2 \times \frac{\text{Preemphasis.length}}{\text{samplePerFrame}-1} \quad (4)$$

Data *pre-emphasis* merupakan data sinyal suara yang dipecah menjadi sebanyak *frame* dan setiap *frame* terdiri dari data sebanyak *sample per frame*. Data yang diuji coba pada penelitian ini menghasilkan panjang *pre-emphasis* sebanyak 15664, yang merupakan data suara yang memiliki *voiced* dan nilai *sample per frame* 512.



Gambar 6. Hasil Framing

Dari nilai tersebut maka jumlah *frame* yang akan dihasilkan sebanyak 60 dimana setiap *frame* terdapat 512 sinyal suara. Gambar 6 berikut menunjukkan hasil grafik sinyal sesudah dipecah menjadi beberapa bagian.

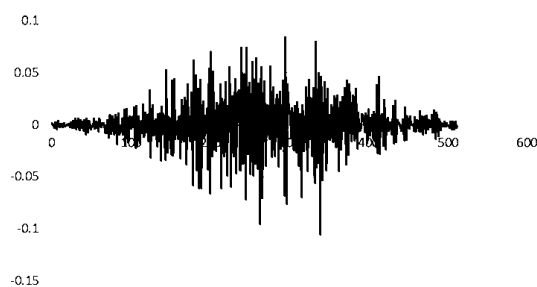
Untuk mengurangi efek diskontinuitas pada ujung-ujung *frame* yang dihasilkan oleh proses *frame blocking*, maka setiap *frame* akan dilakukan proses *windowing* dengan persamaan(5) dan persamaan(6) berikut ini.

$$x(n) = x_i(n)w(n) \quad n = 0, 1, \dots, N - 1 \quad (5)$$

dimana $x(n)$ adalah nilai sampel sinyal hasil *windowing*, $x_i(n)$ adalah nilai sampel dari *frame signal* ke-1, $w(n)$ adalah fungsi *windows* dan N adalah jumlah sampel pada setiap *frame*. Penelitian ini menggunakan *hamming window* sebagai fungsi *windows* ($w(n)$) yang ditunjukkan pada persamaan (6) berikut.

$$w(n) = 0.54 - 0.46 \cos \frac{2\pi n}{M-1} \quad (6)$$

pada persamaan (6) nilai $n > 0$ dan $n < M - 1$, dimana M adalah panjang *frame*. Gambar 7 grafik sinyal sesudah dilakukan proses *windowing*.



Gambar 7. Hasil Windowing

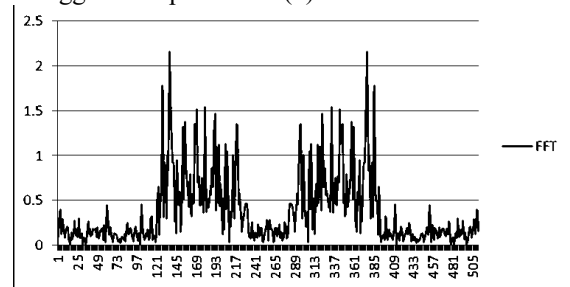
Langkah selanjutnya sinyal suara yang berupa amplitudo yang merupakan domain waktu akan diubah menjadi domain frekuensi dengan FFT untuk mempermudah dalam proses pengolahan sinyal.

FFT merupakan teknik perhitungan cepat dari DFT dengan memanfaatkan sifat periodikal dari *transformasi fourier* dengan menggunakan persamaan (7).

$$S[k] = \sum_{n=0}^{N-1} s[n]e^{-j2\pi nk/N}, \quad 0 \leq k \leq N - 1 \quad (7)$$

pada persamaan (7), N adalah jumlah sampel yang akan diproses, $S[k]$ adalah nilai sampel sinyal dan K adalah variabel frekuensi *discrete*. Untuk sampel yang sama dengan proses sebelumnya, Gambar 8 berikut menunjukkan grafik FFT.

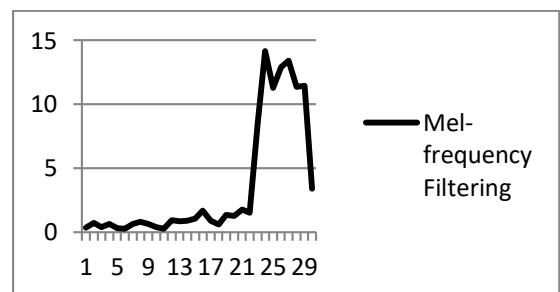
Selanjutnya frekuensi akan dihitung energinya dengan *mel-frequency filtering*. Untuk mengetahui ukuran energi dari setiap *frekuensi band* dengan number *mel-filters* = 30 dihitung dengan menggunakan persamaan(8).



Gambar 8. Hasil FFT

$$Y[i] = \sum_{j=1}^N S[j]H_1[j] \quad (8)$$

Pada persamaan (8), N adalah jumlah *magnitude spectrum*, $S[j]$ adalah *magnitude spectrum* pada frekuensi j , $H[j]$ adalah koefisien *filterbank* pada frekuensi j ($1 \leq i \leq M$) dan M adalah jumlah channel dalam *filterbank*. Gambar 9 berikut menunjukkan contoh hasil grafik *mel-frequency filtering* pada penelitian ini.



Gambar 9. Mel-frequency Filtering

Proses terakhir dari MFCC adalah DCT yang berfungsi untuk mendapat *mel-cepstrum*. Hasil dari DCT adalah fitur-fitur yang digunakan untuk melakukan proses analisa terhadap pengenalan suara. Proses ini dihitung dengan menggunakan persamaan (9).

$$C_n = \sum_{k=1}^K (\log S_k) \cos \left[n \left(k - \frac{1}{2} \right) \frac{\pi}{K} \right]; \quad n = 1, 2, \dots, K \quad (9)$$

pada persamaan(9), S_k adalah keluaran dari proses *filterbank* pada index k dan K adalah jumlah koefisien yang diharapkan. Proses DCT menghasilkan *mel-cepstrums*, dengan jumlah koefisien = 12 pada setiap *frame*.

Selain menghasilkan *mel-cepstrums*, MFCC juga menghasilkan *log energy* untuk digunakan

dalam proses HMM. *Log energy* ini merupakan salah satu cara untuk menambah nilai koefisien yang dihasilkan dari *mel-cepstrums*, nilai tersebut merupakan *log energy signal*. Ini berarti pada setiap *frame* terdapat nilai energi yang ditambahkan. Log energi dapat dinyatakan dengan persamaan(10) berikut :

$$E = \log \sum_{n=1}^N s(n)^2 \quad (10)$$

dimana $s(n)$ adalah sinyal suara didalam *frame* dan N adalah panjang *frame*. Sinyal suara yang digunakan pada log energi adalah sinyal suara yang belum dilakukan operasi *pre-emphasis*. Kemudian dilakukan operasi *frame blocking* untuk membagi sinyal suara menjadi *frame-frame* kecil.

Dalam penelitian ini nilai *log energy* untuk *frame* 1 adalah 4.304689. Untuk meningkatkan akurasi dari fitur MFCC perlu dilakukan juga kombinasi terhadap *mel cepstrals* yang dihasilkan yaitu :

1. 12 MFCC *Feature* yang merupakan hasil *mel-cepstrum*
2. 12 Δ MFCC *Feature* merupakan turunan pertama dari MFCC / hasil *mel-cepstrum*.

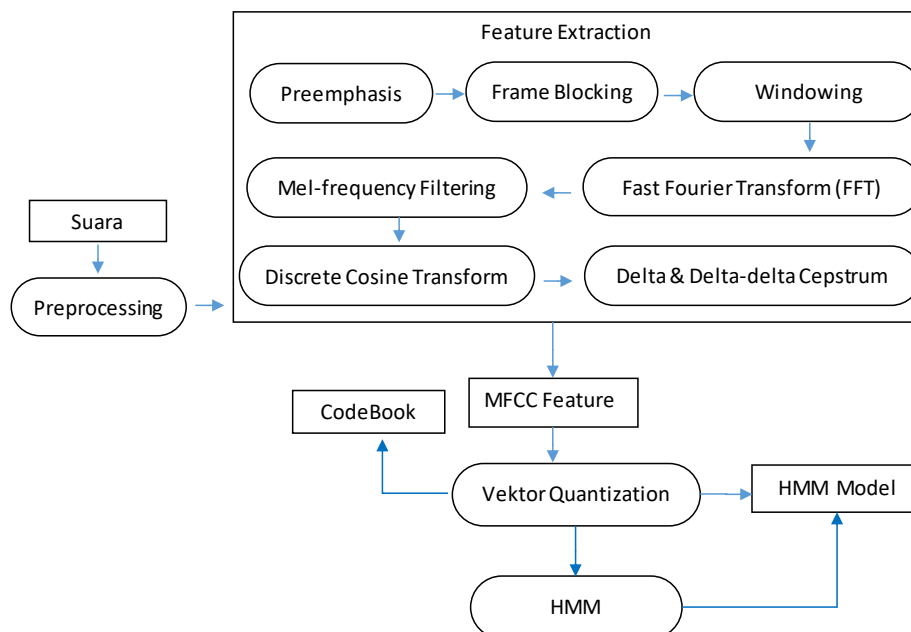
$$d(k) = \frac{\sum_{t=1}^T t * (C(k+t) - C(k-t))}{2 \sum_{t=1}^T t^2}, 1 \leq k \leq N \quad (11)$$

dimana $C(k)$ adalah koefisien cepstral, N adalah jumlah koefisien cepstrum, dan T adalah konstanta. Nilai T yang biasa digunakan adalah 2. Sedangkan *delta-delta cepstrum* didapat dari turunan pertama *delta cepstrum*.

3. 12 $\Delta\Delta$ MFCC *Feature* Merupakan turunan pertama dari 12 Δ MFCC / hasil *mel cepstrum*
4. 1 *Energy* : Nilai *log energy* dari suara
5. 1 Δ *Energy* : Turunan pertama dari nilai 1 *Energy*.
6. 1 $\Delta\Delta$ *Energy* : Turunan kedua dari nilai 1 *Energy*.

2.3. Training

Input proses ini adalah MFCC *feature* termasuk di dalamnya nilai energi. Proses *training* selanjutnya yaitu pembentukan *codebook* untuk proses *vektor quantization* (Mustofa, 2007) dan HMM *model* melalui pemodelan HMM (Prasetyo, Kurniawan & Ichsan, 2017) yang nantinya akan digunakan untuk melakukan pengenalan pola suara. Proses *training* dapat dilihat pada Gambar 10 berikut.



Gambar 10. Blok Diagram Proses Training

Konsep dasar dari *vector quantization* (VQ) adalah memetakan vektor-vektor dari jumlah besar menjadi jumlah tertentu. Dengan bantuan K-Means dibentuk 256 *cluster centroid* dari masing-masing kluster disebut *codeword*. Kumpulan dari semua *codeword* inilah yang disebut *codebook*. Pengurangan jumlah vektor ini akan mempercepat proses saat verifikasi, dimana vektor-vektor data uji tidak akan dibandingkan dengan semua vektor data latih yang jumlahnya sangat banyak, melainkan hanya dengan *centroid*-nya saja yang jumlahnya lebih

sedikit. *Centroid/codeword* merupakan fitur ciri yang kemudian digunakan sebagai input untuk membuat model HMM, yang hasil parameternya nantinya digunakan pada proses pengenalan suara menggunakan Algoritma Viterbi. Dengan menggunakan Algoritma Baum-Welch (Fuentes, Javier & Torres, 2013) vektor suara diolah untuk membentuk HMM model. HMM *model* terdiri dari Vektor π , Matrik A dan Matrik B yang dibentuk dengan persamaan(12) berikut.

$$\pi_i = \gamma_t(i); a_{ij} = \frac{\sum_{t=1}^{T-1} \varepsilon_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}; b_j(k) = \frac{\sum_{t=1}^{T-1} \gamma_t(j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \quad (12)$$

Pada persamaan(12), a_{ij} menyatakan banyaknya transisi yang diharapkan dari state i menuju state j dibagi banyaknya transisi dari state i . $b_j(k)$ menyatakan banyaknya waktu yang diharapkan berada pada state j dan mengamati simbol v_k dibagi banyaknya waktu yang diharapkan berada pada state j . Model awal $\lambda = (A, B, \pi)$ mendefinisikan titik kritis pada fungsi likelihood, jika $\lambda = \lambda$ atau lebih mendekati model λ dalam mewakili peluang deretan observasi, $P(O|\lambda) > P(O|\lambda)$. Maka, dapat dikatakan proses reestimasi telah menemukan model yang lebih baik dibandingkan dengan model awal. Hasil akhir dari prosedur reestimasi disebut maximum likelihood dari HMM.

2.4. Testing

Seperti yang dijelaskan sebelumnya, bahwa proses testing juga melibatkan MFCC untuk ekstraksi fitur. Setelah fitur suara diinputkan, suara juga akan diekstraksi oleh MFCC untuk dicari fiturnya. Fitur ini akan dicari *centroid*-nya di codebook dengan menggunakan *Vector Quantization*(VQ). Setelah itu dengan *Algoritma Viterbi* (Fuentes, Javier & Torres, 2013) dicari probabilitas kemiripan (likelihood) pada setiap model di HMM model. Berikut langkah-langkah dan hasil perhitungannya:

1. Inisialisasi untuk menentukan nilai maksimum awal dalam penentuan proses Viterbi dengan menggunakan persamaan(13) berikut.

$$\delta_1(i) = \pi_i b_i(o_1), 1 \leq i \leq N; \psi_1(i) = 0 \quad (13)$$

2. Rekursif, melakukan perhitungan nilai maksimum dengan data yang lain dengan menggunakan persamaan(14)

$$\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_j(o_t), 2 \leq t \leq T; 1 \leq j \leq N \quad (14)$$

3. Menentukan nilai *likelihood maximum* dari δ_T

3. HASIL DAN PEMBAHASAN

Setelah melakukan pembuatan aplikasi *e-learning* dengan menggunakan MFCC dan HMM, maka selanjutnya dilakukan uji coba untuk mengetahui tingkat keberhasilan sistem yang telah dibuat. Uji coba dilakukan dengan menggunakan suara dari 10 anak usia dini. Suara yang dipilih hanya suara yang dinilai benar dalam pengucapannya. Setiap anak diminta menyebut sejumlah nama objek dalam Bahasa Madura untuk kemudian direkam dan dijadikan sebagai data *training*.

	Anak									
	1	2	3	4	5	6	7	8	9	10
Objek 1										
Objek 2										
Objek 3										
Objek 4										
Objek 5										

Ket : Training
 Testing

Gambar 11. Ilustrasi Skenario Percobaan I

Uji coba dilakukan dalam dua tahap percobaan yaitu uji coba dengan menggunakan satu model dan multi model. Setiap percobaan ini dilakukan dalam dua bentuk yaitu percobaan I dan percobaan II. Percobaan I dilakukan dengan menjadikan suara dari 8 anak sebagai data training dan suara 2 anak lainnya sebagai data testing. Percobaan II dilakukan dengan menjadikan beberapa suara dari 10 anak sebagai data *training* dan beberapa suara lainnya sebagai data *testing*. Gambar 11 menunjukkan skenario uji coba pada percobaan I dan Gambar 12 menunjukkan skenario uji coba pada percobaan II.

	Anak									
	1	2	3	4	5	6	7	8	9	10
Objek 1										
Objek 2										
Objek 3										
Objek 4										
Objek 5										

Ket : Training
 Testing

Gambar 12. Ilustrasi Skenario Percobaan II

Setelah dilakukan pengujian sistem di atas, diketahui bahwa pengujian dengan multi model lebih baik dari pada satu model. Dengan percobaan multi model, untuk satu kelas memiliki banyak model sehingga mengurangi resiko kesalahan untuk beberapa gaya pengucapan suara yang berbeda. Hal ini karena masih terdapat data suara yang memiliki perbedaan pengucapan pada nama objek yang sama seperti pada objek “kalambih (baju)”, ada yang mengucapkan “kalambih” dan ada juga yang mengucapkan “kelambih” dimana “e” dibaca seperti dalam kata “ke” pada Bahasa Indonesia. Meskipun pengucapan ini berbeda namun masih bisa dibenarkan karena adanya perbedaan kebiasaan pengucapan (dialek) dalam lingkungan. Tabel 1 menunjukkan perbandingan hasil antara percobaan satu model dan multi model.

Pengujian sistem dengan satu model, hasil akurasi percobaan II lebih baik dari percobaan I. Pada percobaan I, adanya suara yang tidak valid (kurang benar pengucapannya) mengakibatkan model suara untuk suatu objek menjadi berubah dari seharusnya sehingga jika ada suara yang dicobakan untuk objek tersebut memiliki kemungkinan lebih besar tidak

tertebak dibandingkan dengan uji coba pada percobaan II.

Tabel 1. Hasil Pengujian Sistem

No	Percobaan	Rata-Rata Perhitungan		
		Benar	Jumlah Percobaan	Rata-rata Akurasi
1	Satu Model Percobaan.I	47	64	73 %
2	Satu Model Percobaan II	48	64	75 %
3	Multi Model Percobaan I	52	64	81 %
4	Multi Model Percobaan II	51	64	80 %

Hal ini karena pada percobaan II, data training merupakan suara acak dari 10 anak, sehingga tidak semua data training memiliki unsur dari suara anak yang dianggap tidak valid, sehingga beberapa suara tetap memiliki model suara yang valid. Oleh karena itu percobaan multi model perlu dilakukan sehingga, suara yang tidak valid tidak akan merusak model suara yang valid, karena setiap suara anak dibuatkan model untuk masing-masing objek. Terbukti pada percobaan yang dilakukan, rata-rata akurasi untuk multi kelas lebih tinggi dari pada satu kelas.

Percobaan II untuk multi model mengalami penurunan akurasi dibandingkan percobaan I untuk multi model, karena kecenderungan objek yang tertebak adalah objek yang namanya menyerupai nama objek yang sedang ditebak dan data training dari objek tertebak menggunakan suara anak yang sedang dites. Sebagai contoh, seorang anak bernama Bilqis suaranya digunakan untuk data training pada objek “Manok (Burung)”, kemudian suara anak ini dijadikan data tes untuk “Jukok (Ikan)”. Pada kasus ini ada kemungkinan objek yang tertebak adalah “Manok” karena penyebutan nama objek hampir sama dan didukung oleh suara training yang digunakan adalah suara anak tersebut (Bilqis), sehingga nilai *likelihood maximum* berada pada objek “Manok” untuk suara Bilqis.

4. KESIMPULAN

Aplikasi *e-learning*/pembelajaran untuk anak usia dini dengan memanfaatkan pengenalan pola suara dapat digunakan untuk pembelajaran Bahasa Madura yang memiliki banyak bentuk pengucapan dalam intonasinya. Untuk hasil terbaik, data latih yang digunakan harus merupakan data latih dengan *noise* rendah dan pengucapan yang benar sehingga hasilnya sesuai dengan yang diharapkan. Pengenalan pola suara dengan MFCC dan HMM mencapai akurasi terbaik dengan multi model karena dapat menampung sejumlah model suara untuk objek yang berbeda sehingga meningkatkan kemungkinan suara dapat dikenali dengan benar sekalipun pengucapan sedikit berbeda. Dengan mengumpulkan data suara yang beragam dan benar dalam pengucapannya, maka

penelitian selanjutnya adalah pemanfaatan pola suara untuk memperkenalkan Bahasa Madura sesuai dengan tingkatan tutur. Cara penulisan dan cara membaca kata Bahasa Madura yang cukup sulit juga dapat dijadikan sebagai objek penelitian lainnya.

UCAPAN TERIMA KASIH

Terima kasih kepada Direktorat Riset dan Pengabdian kepada Masyarakat khususnya Direktorat Jenderal Penguatan Riset dan Pengembangan yang telah memberikan kontribusi berupa bantuan dana sehingga penelitian ini dapat terlaksana dengan baik.

DAFTAR PUSTAKA

- ABDULLA, W. H., CHOW, D., & SIN, G., 2003. Cross-words reference template for DTW-based speech recognition systems. In TENCON 2003. Conference on Convergent Technologies for the Asia-Pacific Region, Vol.4, pp. 1576-1579.
- CANTIKA, P.D., B. HERMANTO., 2018. Rancang Bangun Aplikasi *E-learning* Untuk Pembelajaran Agama Islam Berbasis Android (Studi Kasus MIN 6 Bandar Lampung). Jurnal Komputasi, 6(1), pp. 25-32.
- DESHMUKH S.D., BACHUTE, M.R., 2013. Automatic Speech and Speaker Recognition by MFCC, HMM and Vector Quantization. International Journal of Engineering and Innovative Technology (IJEIT), 3(1), pp.109-113.
- DEWI, N.P., I. LISTIOWARNI., 2019. Implementasi Game Based Learning pada Pembelajaran Bahasa Inggris. Jurnal Resti, 3(2), pp.124-130.
- FUENTES, R., JAVIER, L., TORRES, M., 2003. Comparative Study of the Baum-Welch and Viterbi Training Algorithms Applied to Read and Spontaneous Speech Recognition, pp 847 – 857.
- GUPTA, S., JAAFAR, J., AHMAD, W.F.W., & BANSAL, A., 2013. Feature Extraction Using MFCC. Signal & Image Processing : An International Journal (SIPIJ), 4(4), pp. 101-108.
- IBRAHIM, Y.A., ODIKETA, J.C., IBIYEMI, T.S., 2017. Preprocessing Technique In Automatic Speech Recognition for Human Computer Interaction: An Overview. Annals Computer Science Series, 15(1), pp.186-191.
- IQBAL, S., MAHBOOB, T., KHIYAL, M.S.H., 2011. Voice Recognition using HMM with MFCC for Secure ATM. IJCSI International Journal of Computer Science Issues, 8(3), pp. 297-303.
- MEHTA, L. R., DABHADE, A. S., 2013. Comparison Study of MFCC and LPC for Marathi Isolated Word. Int. J. Adv. Res. Electr. Electron. Instrum. Eng., 2(6), pp. 2133-2139.
- MON, S.M., TUN, H.M., 2015. Speech-To-Text Conversion STT System Using Hidden Markov

- Model HMM. International Journal of Scientific & Technology Research, 4(6), pp. 349-352(4).
- MUSTOFA, A., 2007. Sistem Pengenalan Penutur dengan Metode Mel-frequency Wrapping. Jurnal Teknik Elektro, 7(2), pp. 88 – 96.
- NAZIYA, S., DESHMUKH, S.D., 2015. Speech Recognition System – A Review_IOSR Journal of Computer Engineering (IOSR-JCE), 18(4), pp.01-09.
- PANCHAL, J.R., 2010. Linear Predictive Coding For Speech Signal Feature Extraction. Thesis. California State University, Sacramento.
- PERMANA, A.F., 2011. Implementasi Hidden Markov Model Untuk Aplikasi Pengenalan Ucapan Sebagai Kendali Gerak Robot Mobil. Tugas Akhir. Institut Pertanian Bogor.
- PRASETIO, B.H., KURNIAWAN, W., ICHSAN, M.H.H., 2017. Pengenalan Emosi Berdasarkan Suara menggunakan Algoritma HMM. Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK), 4(3), pp.168-172.
- SAHA, G., CHAKROBORTY, S., SENAPATI, S., 2014. A New Silence Removal and Endpoint Detection Algorithm for Speech and Speaker Recognition Applications, IJIGSP, pp. 1-5.
- SANJAYA, W.S.M., SALLEH, Z., 2014. Implementasi Pengenalan Pola Suara Menggunakan Mel_Frequency Cepstrum Coefficients (MFCC) dan Adaptive Neuro-Fuzzy Inferense System (ANFIS) sebagai Kontrol Lampu Otomatis, Al-Hazen Journal of Physics, 1(1), pp.44-54.
- SANJAYA, W.S.M., 2017. Robot Cerdas Berbasis Speech Recognition. Yogyakarta: ANDI.
- SOFYAN, A., 2007. Dialek dan Tingkat Tutur dalam Bahasa Madura. Jurnal Medan Bahasa, Vol.1 No.1.
- SOFYAN, A., 2017. Tata Bahasa Bahasa Madura. Sidoarjo: Bahasa Surabaya.
- TIWARI, V., 2010. MFCC and Its Applications in Speaker Recognition. Int. J. Emerg. Technol., 1(1), pp. 19–22.
- YUNDATAMA, Y., PARNO. 2015. *E-learning* English for Children Aplikasi Mobile Mengenal Alat Transportasi Berbasis Android. Prosiding PESAT (Psikologi, Ekonomi, Sastra, Arsitektur & Teknik Sipil), Vol. 6, pp.134-141.

Halaman ini sengaja dikosongkan