Jurnal Teknologi Informasi dan Komunikasi
Vol: 13 No 02 2022 **E-ISSN: 2477-3255**
Diterima Redaksi: 24-07-2022 | Revisi: 26-09-2022 | Diterbitkan: 30-11-2022

Speech Recognition for English Sentences with Malay Accent

Keumala Anggraini[1], Lucky Lhaura Van FC[2], Yuvi Darmayunata[3]
[1,2,3]Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Lancang Kuning
[1,2,3]Jl. Yos Sudarso KM. 8 Rumbai, Pekanbaru, Riau, telp. 0811 753 2015
e-mail: [1]keumala@unilak.ac.id, [2]lucky@unilak.ac.id, [3]yuvidarmayunata@gmail.com

Abstract

Some countries conduct speech research using accents. One language that has a different accent is English. English is an international language that is often used to communicate with citizens of other countries. For beginner, many difficulty to translate English with accent, include malay accent. This study performs speech recognition using English and Riau Malay accents. This research use Google Recognizer to cut words in sentences, Mel Frequency Cepstral Coefficient for feature extraction, and Hidden Markov Model for classification. The accuracy of this research is 94.02%.

Keywords: *Speech, Accent, Google Recognizer, Mel Frequency Cepstral Coefficient, Hidden Markov Model*

Pengenalan Suara Kalimat Bahasa Inggris dengan Aksen Melayu

Abstrak

Beberapa negara melakukan penelitian tuturan menggunakan aksen. Salah satu bahasa yang mempunyai aksen berbeda adalah Bahasa Inggris. Bahasa Inggris merupakan bahasa internasional yang sering digunakan untuk berkomunikasi dengan warga negara lain. Untuk pemula, terdapat banyak kesulitan dalam menerjemahkan bahasa inggris yang dilafalkan menggunakan aksen, termasuk aksen Bahasa Melayu. Oleh karena kesulitan itu, penelitian ini melakukan speech recognition menggunakan Bahasa Inggris dan aksen Melayu Riau. Penelitian ini menggunakan Google Recognizer untuk memotong kata dalam kalimat, Mel Frequency Cepstral Coefficient untuk ekstraksi fitur, dan Hidden Markov Model untuk klasifikasi. Akurasi yang dihasilkan pada penelitian ini sebesar 94,02%.

Kata kunci: *Tutur, Aksen, Google Recognizer, Mel Frequency Cepstral Coefficient, Hidden Markov Model*

1. Pendahuluan

Berbicara atau dalam bahasa Inggris *speech* selalu menjadi komunikasi utama antara orang-orang yang berkomunikasi. Banyak informasi yang disampaikan secara langsung melalui *speech* atau pembicaraan dari pembicara ke pendengar dengan waktu yang singkat. Perkembangan zaman memungkinkan komunikasi dapat dilakukan oleh manusia dan mesin

seperti *Google Assistant, Apple's Siri*, *Alexa*, dan lainnya [1]. Pengenalan pembicaraan biasanya lebih dikenal dengan *speech recognition*. *Speech recognition* merupakan suatu proses yang digunakan untuk mengenali suatu pembicaraan yang diutarakan oleh pembicara. *Speech recognition* sudah dikembangkan dari tahun 1950. Banyak contoh tentang *speech recognition* pada kehidupan sehari-hari, yaitu pemesanan tiket menggunakan suara, *home automation system*, *Google voice application*, dan lainnya [2].

Identifikasi dan klasifikasi logat atau aksen merupakan bagian penting dari sistem speech recognition. Dengan meningkatnya penggunaan pendamping suara pada mesin, maka sistem speech recognition pada bagian logat atau aksen juga meningkat. Sebagai contoh ketika berbicara bahasa Inggris, maka terdapat aksen Amerika dan aksen British [3]. Tidak hanya Bahasa Inggris menggunakan aksen Amerika dan aksen British. Karena bahasa Inggris merupakan bahasa Internasional yang digunakan untuk berkomunikasi, oleh karena itu Bahasa Inggris banyak mempunyai aksen atau logat dari berbagai negara didunia, sebagai contoh menggunakan aksen India, Korea, Italia dan lainnya [4].

Banyak penelitian-penelitian mengenai *speech recognition* menggunakan aksen dari beberapa negara di dunia. Salah satu diantaranya berjudul *Deep Learning Approach to Accent Classification*. Penelitian ini melakukan pengenalan Bahasa Inggris menggunakan *Deep Learning*. Penelitian ini menggunakan dataset yang diambil dari Wildcat Corpus yang berisi percakapan dua orang dalam Bahasa Inggris, satu orang adalah *native speaker* dan lainnya adalah *non-native speaker*. Wildcat corpus terdiri dari *speaker* asli dari Cina, Inggris, dan Korea. Tahapan yang dilakukan pada penelitian ini dimulai dari tahap preproses. Tahap preproses digunakan untuk mengelompokkan ucapan kata pada setiap rekaman audio dan mengekstraksinya ke sinyal audio. Selanjutnya melakukan ekstraksi fitur menggunakan MFCC (Mel Frequency Cepstral Coefficient). Metode klasifikasi pada penelitian ini adalah CNN (Connvolutional Neural Network. Akurasi yang dihasilkan sebesar 88% [5].

Penelitian lain tentang *speech recognition* berjudul "*Javanese Gender Speech Recognition Based on Machine Learning Using Random Forest and Neural Network*" menjelaskan mengenai pengenalan suara dalam kata Bahasa Jawa yang dituturkan oleh 5 orang pria dan 5 orang wanita. Dataset yang digunakan pada penelitian ini adalah rekaman suara yang merekam kata "makan", "minum" dan "tidur" dalam Bahasa Jawa. Ekstraksi Fitur yang digunakan pada penelitian ini adalah MFCC. Model klasifikasi yang digunakan adalah *Neural Network* dan *Random Forest*. Akurasi yang dihasilkan sebesar 92,2 % untuk model *neural network* dan 91,3% untuk model *random forest* [6].

Penelitian [7] mencoba untuk melakukan pengenalan Bahasa Spanyol menggunakan *automatic speech recognition* menggunakan HMM atau *Hidden Markov Model*. Dataset yang digunakan adalah 1621 rekaman suara yang telah diberi anotasi atau label. Selanjutnya adalah tahapan konversi audio set menjadi *phoneme*. *Phoneme* yang dihasilkan adalah *uppercase vowel* {A, E, I, O, U}, *semivowel*, *short pause*, dan *silenece*. Selanjutnya adalah klasifikasi menggunakan HMM. Akurasi yang dihasilkan pada penelitian ini sebesar 77,91%.

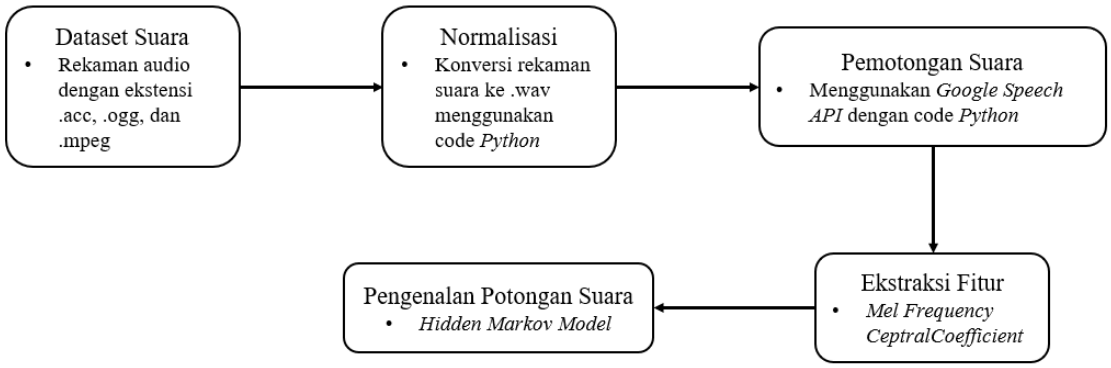
Penelitian [5] dan penelitian [6] menunjukkan bahwa pengenalan suara menggunakan ekstraksi fitur MFCC menghasilkan akurasi lebih dari 85%. Penelitian [7] menunjukkan bahwa pengenalan suara menggunakan model klasifikasi HMM mendapatkan akurasi diatas 75%. Penelitian [5] juga menunjukkan hasil preproses yang dapat dilakukan dengan menggunakan sinyal suara yang dipotong sesuai dengan *time window*. Namun, praproses ini sulit dilakukan jika pembicara atau *speaker* mempunyai nada pelafalan aksen yang cepat dan lambat seperti aksen pada daerah yang terdapat pada Indonesia. Dan penelitian *speech recognition* menggunakan aksen atau logat melayu Riau belum dilakukan. Oleh karena itu, penelitian ini mencoba untuk melakukan pengenalan potongan suara Bahasa Inggris dengan aksen atau logat melayu Riau. Penelitian ini menggunakan dataset yang berupa rekaman suara beberapa orang yang mengucapkan kalimat Bahasa Inggris menggunakan aksen atau logat melayu Riau. Penelitian ini menggunakan Google Speech API sebagai preproses, MFCC sebagai ekstraksi fitur, dan pengenalan potongan kata menggunakan HMM.

https://doi.org/10.31849/digitalzone.v13i2.10579

Berbeda dengan penelitian sebelumnya, yaitu *speech recognition* lebih cenderung untuk mengenali rekaman yang berisi audio seseorang yang melafalkan suatu kata atau rekaman audio seseorang yang melafalkan *viseme-phoneme*. Penelitian ini menggunakan rekaman audio seseorang yang mengucapkan kalimat dalam Bahasa Inggris, yang kemudian dilakukan proses pemotongan suara. Sehingga, penelitian ini lebih berfokus kepada pemotongan suara pada rekaman audio. Selanjutnya, hasil pemotongan akan dikenali menggunakan *Hidden Markov Model.*

2. Metode Penelitian

Metode penelitian yang digunapan pada penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Metode Penelitian

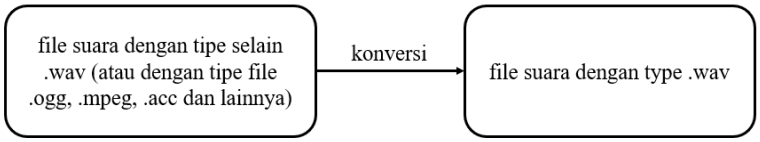
Dataset yang digunakan pada penelitian ini adalah rekaman audio kalimat dalam bahasa Inggris yang diucapkan oleh 4 (empat) orang yang terdiri dari 2 (dua) orang wanita dan 2 (dua) orang pria. Penutur kalimat adalah mahasiswa Universitas Lancang Kuning yang berdomisili atau lahir di Riau. Terdapat 4 (empat) kalimat dalam Bahasa Inggris yang sering digunakan dalam kehidupan sehari-hari atau kalimat yang mudah dipahami oleh mahasiswa Univeritas Lancang Kuning. Setiap penutur akan merekam kalimat sebanyak 10 rekaman pada setiap kalimat. Dataset berupa rekaman audio yang direkam pada Universitas Lancang Kuning, Pekanbaru, Riau. Dataset ini selanjutnya akan dinamakan audioset. Daftar kalimat yang digunakan pada penelitian ini dapat dilihat pada Tabel.1. Rekaman audio masih menggunakan berkas audio bertipe .acc, .ogg, dan .mpeg, sehingga rekaman audio harus dinormalisasi untuk diproses ketahap selanjutnya. Tahap selanjutnya adalah tahap pemotongan suara menggunakan *Google Speech API*. Hasil dari pemotongan suara akan melalui tahap ekstraksi fitur menggunakan *Mel Frequency Cepstral Coefficient* dan tahap pengenalan potongan suara menggunakan *Hidden Markov Model*. Proses metode penelitian hingga mendapatkan hasil pengenalan potongan suara pada penelitian ini menggunakan bahasa pemograman *Python* serta *library* yang berkaitan.

Tabel 1. Daftar Dataset

	Kalimat
Kalimat 1	Good morning, how are you today
Kalimat 2	Are you feeling weel?
Kalimat 3	I think you must going to doctor
Kalimat 4	I am eating something fishy

2.1. Normalisasi

Normalisasi adalah tahapan awal yang harus dilakukan pada penelitian ini. Normalisasi berfungsi untuk menyamaratakan tipe berkas dari dataset rekaman suara. Normalisasi pada penelitian ini dapat dilihat pada Gambar 2.



Gambar 2. Normalisasi

Rekaman suara pada dataset masih bertipe berkas audio .ogg, .mpeg, dan .acc. Tipe berkas audio .wav menjadi acuan penyetaraan tipe berkas audio pada penelitian ini. Sehingga audioset harus melalui proses konversi untuk mendapatkan hasil berkas audio .wav. Konversi pada penelitian ini akan menggunakan *library Python pydub AudioSegment*. Tabel 2 merupakan *pseudocode* pada normalisasi.

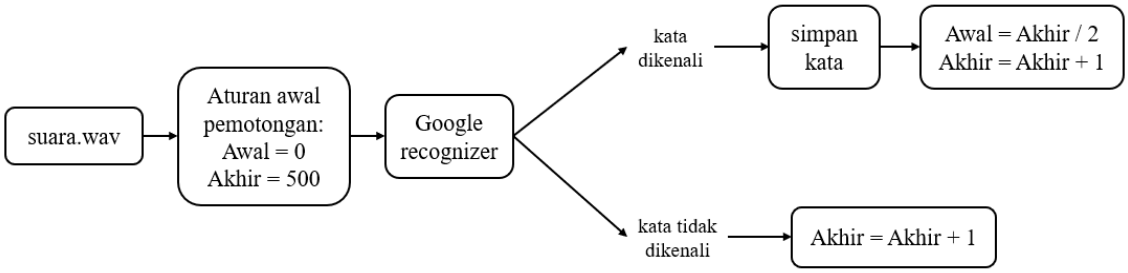
Tabel 2. *Pseudocode* Normalisasi

Program: Normalisasi
Input : audioset.acc or audioset.mpeg or audioset.ogg
Algoritma: read **input** using **AudioSegment** export to .wav

2.2. Preproses

Preproses adalah tahapan selanjutnya setelah tahapan normalisasi dilakukan. Preproses pada penelitian ini berguna untuk memotong rekaman suara yang masih berbentuk kalimat sehingga mendapatkan potongan kata dari kalimat tersebut. Tahapan preproses penelitian ini menggunakan *library Python SpeechRecognitioon* yang didukung oleh *Google Cloud Speech API.*

Google Cloud Speech API merupakan *automatic speech recognition* (ASR) yang menggunakan model pembelajaran *deep neural network*. *Google Cloud Speech API* ini dapat digunakan untuk pencarian menggunakan suara dan penerjemahan bentuk suara ke bentuk teks. *Google Cloud Speech API* juga dapat mengatasi suara-suara yang mempunyai keributan atau *noise*. *Google Cloud Speech API* dapat memahami suara yang menyatakan kalimat dan pengenalan paling baik pada bahasa-bahasa yang baku atau bahasa standarnya [8].



Gambar 3. Preproses

https://doi.org/10.31849/digitalzone.v13i2.10579

Gambar 3 merupakan tahapan preproses pada penelitian ini. Audioset yang telah melalui tahap normalisasi akan masuk ketahap preproses. Pada tahap preproses, audioset yang telah bertipe berkas .wav atau suara.wav akan diberi aturan awal mulainya pemotongan audioset. Aturan awal pada preproses adalah menentukan waktu (*time*) awal pada audio dan waktu (*time*) akhir pada audio. Hal ini dilakukan karena pada audio terdapat parameter waktu (*time*), sehingga ketika pemotongan dilakukan berdasarkan waktu awal dan waktu akhir, *google recognizer* akan melakukan penerjemahan dari waktu awal hingga waktu akhir. Langkah selanjutnya adalah jika kata dikenali oleh *google recognizer*, maka kata yang telah dikenali akan disimpan dan waktu awal akan diganti menjadi pertengahan waktu akhir yang sebelumnya dan waktu akhir akan ditambahkan 1 (satu). Kemudian, jika kata tidak dikenali oleh *google recohnizer*, maka waktu awal akan tetap dan waktu akhir akan ditambahkan dengan 1. Preproses pada penelitian ini menggunakan *library Python SpeechRecognizer google_recognizer*. *Pseudocode* pada preproses penelitian ini dapat dilihat pada Tabel 3.

Tabel 3. *Pseudocode* Preproses

```
Program:
    Preproses

Input:
    audio.wav
    awal = integer
    akhir = integer

Algoritma:
    read audio.wav using AudioSegment
    awal = 0
    akhir = 500

    while audio.wav is opened:
        recognize_google for audio[awal:akhir]

        if audio is recogntzed:
            save audio
            awal = akhir / 2
            akhir = akhir + 1

        else:
            akhir = akhir + 1
```

2.3. Ekstraksi Fitur

Ekstraksi fitur merupakan tahapan untuk mengambil fitur-fitur yang terdapat pada suatu data didalam dataset. Pada penelitian ini menggunakan data suara atau audio, sehingga ekstraksi fitur yang dilakukan harus berkaitan dengan audio. Oleh karena itu, penelitian ini menggunakan ekstraksi fitur MFCC atau *Mel Frequency Cepstral Coefficient*.

MFCC atau *Mel Frequency Cepstral Coefficient* dapat digunakan sebagai fitur yang baik untuk mewakili suara manusia atau musik. MFCC telah terbukti berguna untuk pengenalan suara [9]. Metode MFCC banyak dipilih dikarenakan MFCC mempunyai tingkat akurasi yang baik dalam *speech recognition* [10]. MFCC merupakan metode untuk mengetahui atau mengenali suara yang unik dari manusia [6]. MFCC terdiri dari *preemphasis*, *framing*, *windowing*, *Fast Fourier Transform* (FFT), *Mel Filter Bank*, dan *Discrete Cosine Transform* (DCT). Proses pertama pada metode MFCC adalah *preemphasis*, dimana menghasilkan energi

dalam frekuensi tinggi pada sebelumnya telah dikompresi oleh suara. Kemudian *framing* digunakan untuk memangkas sinyal menjadi bagian kecil. Data audio biasanya berkisar 10-30 ms, karena itu proses analisis sinyal dilakukan pada waktu yang singkat pada *speech recognition*. *Windowing* digunakan untuk menghindari diskontiunitas sinyal yang dihasilkan pada tahap sebelumnya. FFT digunakan untuk mengkonversi sinyal dari bentuk waktu (*time*) ke bentuk frekuensi. *Filterbank* adalah *bandpass filter* yang menindih satu sama lainnya. Tahap terakhir adalah DCT yang menghasilkan koefisien dari MFCC [11]. Ekstraksi fitur menggunakan MFCC pada penelitian ini dapat dilihat pada Gambar 4.

Gambar 4. Ekstraksi Fitur

Tahap ekstraksi fitur dimulai dari data-data audio yang telah dipotong pada tahap sebelumnya. Kemudian data-data audio tersebut akan melalui metode ekstraksi fitur MFCC. Hasil dari ekstraksi fitur MFCC akan disimpan dalam tipe file .npy untuk setiap data-data potongan audionya. Ekstraksi fitur pada penelitian ini menggunakan *library Python librosa* dan *numpy*. *Pseudocode* ekstraksi fitur dapat dilihat pada Tabel 4.

Tabel 4. *Pseudocode* Ekstraksi Fitur

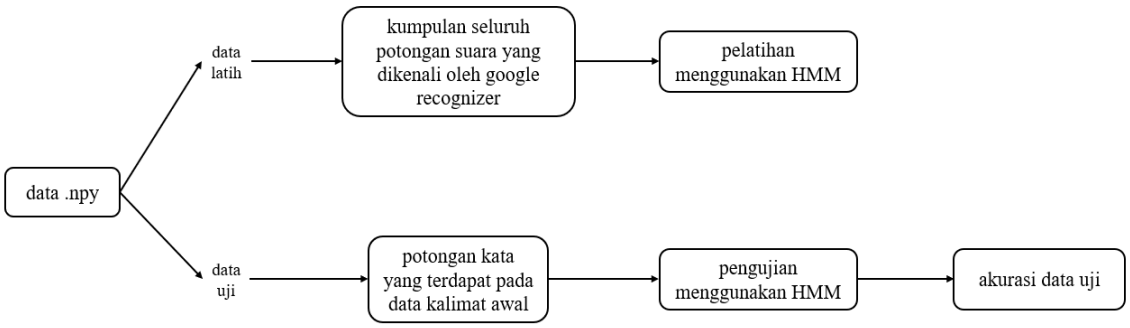
Program: Ekstraksi Fitur
Input: potongan_audio.wav
Algoritma: features = mfcc (potongan audio.wav) save_features to potongan_audio.npy

2.4. Klasifikasi atau Pengenalan

Klasifikasi adalah tahapan lanjutan setelah tahapan ekstraksi fitur telah dilakukan. Klasifikasi yang dilakukan pada penelitian ini adalah melakukan pengenalan pola-pola atau fitur-fitur yang telah diekstrak sebelumnya. Klasifikasi yang dilakukan pada penelitian ini menggunakan metode *Hidden Markov Model* (HMM).

Dalam klasifikasi sinyal suara, HMM digunakan oleh peneliti. Peneliti di CMU dan IBM menggunakan HMM sejak tahun 1970-an. Hal ini dikarenakan, HMM merupakan pendekatan statistik yang cocok untuk pengenalan atau *recognition* [12]. Metode HMM banyak digunakan pada *speech recognition* untuk mengenali emosi dalam ucapan [13]. Pada sistem pengenalan suara, HMM digunakan untuk menggambarkan hubungan transformasi antara keadaan implisit dan deret waktu. HMM adalah model probabilistik yang menggunakan parameter untuk menggambarkan karakteristik statistik dari proses stokastik, dan memiliki kemampuan luar biasa untuk memodelkan deret waktu dinamis [14]. *Hidden Markov Model* merupakan pendekatan statistik umum untuk masalah yang linear yaitu urutan atau deret angka. HMM telah banyak digunakan pada *speech recognition* selama dua puluh tahun. Pada HMM, terdapat kemungkinan untuk menghubungkan teknik-teknik probabilistik dan pengaturan struktur bercelah (*gapped structure arragements*). HMM menawarkan teori untuk penyisipan dan penghapusan bentukan, serta struktur konstan untuk menggabungkan data struktural dan urutan [15]. Klasifikasi menggunakan HMM pada penelitian ini dapat dilihat pada Gambar 5.

https://doi.org/10.31849/digitalzone.v13i2.10579



Gambar 5. Proses Pengenalan Pemotongan Suara

Tahap klasifikasi dimulai dari data yang telah melewati ekstraksi fitur atau data yang telah berbentuk .npy akan dipisah menjadi data latih dan data uji. Pada penelitian ini data latih adalah seluruh data hasil ekstraksi fitur. Sedangkan, data uji adalah data target pemotongan. Data latih dilatih menggunakan metode HMM dan data uji diuji menggunakan HMM. Langkah akhir pada tahap klasifikasi adalah menghitung akurasi dari hasil yang berhasil dikenali. Klasifikasi pada penelitian ini menggunakan library Python hmmlearn. Pseudocode pada klasifikasi penelitian ini dapat dilihat pada Tabel 5.

Tabel 5. *Pseudocode* Klasifikasi

Program:
Klasifikasi

Input:
potongan_suara.py

Algoritma:
read library **hmm** from **hmmlearn**
model = hmm
fit potongan_suara.py to **model**
compile potongan_suara.py
get **score** from testing data

3. Hasil dan Pembahasan

Terdapat beberapa hasil yang ditemukan pada penelitian ini, yaitu hasil pemotongan dan hasil pengenalan menggunakan HMM.

1. Hasil pemotongan
   Pemotongan dataset suara yang berbentuk kalimat diubah menjadi berbentuk kata menggunakan *google recognizer*. Hasil pemotongan dapat dilihat pada Tabel 6.

Tabel 6. Perbandingan Potongan Kata dan Pemotongan *Google Recognizer*

Dataset Kalimat	Target Pemotongan	Hasil Pemotongan *Google Recognizer*
Good morning, how are you today	Good, Morning, How, Are, You, Today	Good, Good morning, Good more, Could more, Goos morn, My, More, Morning, Running height, Morning how are you, I mean how are you today, Funny how are you today, How are you today, Are you today, 8 today, play today, free today, it today, date today, today, dan lainnya

Are you feeling well	Are, You, Feeling, Well	AR, Are you, Are you free, Are you sure, You feel, Youfit, You feeling, You feel me, Eufy, When will, Feeling, Living will, Living well, Linway, Lin weld, , Feeling weel, Linguee, Lingual, Doing well, Baleen Whale, dan lainnya.
I think you must going to doctor	I, Think, You, Must, Going, To, Doctor	It, Fighting, Knitting, Eating, 18, Waiting, Meeting, Yuma, Yuma's, You must, You must go, Yuma's going, You must going, You must be going, Umass going, Mass going, Ask going, Sing, Going, Going to, Playing the doctor, Playing to doctor, Lying to doctor, Going to doctor, The Doctor, To Doctor, Doctor, dan lainnya.
I am eating sonething fishy	I, Am, Eating, Something, Fishy	It, It did, It did some, It did something, It's, Did something, Sissy, Sing ABC, Something, Something fishy PC, Christy, Chrissy, CC, dan lainnya.

Temuan dari hasil pemotongan dengan metode *google recognizer* pada penelitian ini kurang tepat. Pada analisa yang dilakukan pada metode *google recognizer* ditemukan bahwa terdapat beberapa dua atau lebih kata yang dipotong dapat menjadi satu kata yang berbeda makna dengan kata yang digabungkan. Bahkan dua pelafalan dalam satu kata yang seharusnya belum selesai dikenali dapat menjadi kata yang berbeda makna. Kebanyakan kata tunggal atau kata yang mempunyai satu pelafalan tidak dikenali pada *google recognizer* pada saat dilakukan pemotongan kata.

2. Hasil pengenalan menggunakan HMM

Pengenalan dataset suara pada penelitian ini menggunakan metode HMM. Terdapat pengukuran pada hasil penelitian ini. Terdepat dua pengukuran akurasi pada penelitian ini dengan menggunakan rumus berikut.

Rumus 1. Pengukuran Pertama

$$Akurasi = \left(\frac{Seluruh\ Jumlah\ Kata\ yang\ dikenali}{Seluruh\ Jumlah\ Kata\ pada\ Pemotongan}\right) \times 100\%$$

Pengukuran pertama adalah pengukuran akurasi untuk seluruh data uji pada penelitian ini. Akurasi untuk pengukuran pertama atau akurasi untuk seluruh data uji adalah 94,02%.

Pengukuran kedua adalah pengukuran untuk setiap kata sesuai target yang diharapkan. Akurasi pengukuran kedua dapat dilihat pada Tabel 7.

Tabel 7. Akurasi Pengenalan kata

Kata	Jumlah Kata dalam Pemotongan	Jumlah Dikenali	Akurasi (%)
Good	38	38	100
Morning	39	39	100
How	6	6	100
Are	0	0	-
You	3	0	0
Today	40	40	100

Feeling	2	0	0
Well	10	10	100
I	4	4	100
Think	1	0	0
Must	0	0	-
Going	7	7	100
To	0	0	-
Doctor	11	11	100
Am	0	0	-
Eating	11	11	100
Something	11	7	63
Fishy	1	0	0

Temuan dari hasil pengenalan menggunakan HMM, terlihat bahwa akurasi mencapai 94,02%. Pada analisa yang dilakukan pada metode klasifikasi HMM, tingginya akurasi dikarenakan data uji pada penelitian ini merupakan data latih yang sesuai dengan target pemotongan suara pada Tabel 6. Selain itu, perhitungan akurasi didapatkan dari perhitungan hasil yang dikenali yang kemudian dibagi dengan jumlah data uji, sehingga target pemotongan kata yang tidak ada pada data uji akan dianggap sebagai data uji.

4. Kesimpulan

Penelitian ini telah melakukan pengenalan potongan suara menggunakan google recognizer sebagai metode pemotongan suara, MFCC sebagai metode ekstraksi fitur, dan HMM sebagai metode klasifikasi. Terdapat beberapa temuan dan analisa pada penelitian ini yang telah dijabarkan pada pembahasan. Akurasi penelitian ini lebih baik dari penelitian ini lebih baik dari penelitian-penelitian sebelumnya yang telah dijabarkan. Akurasi pada penelitian ini adalah 94,02%, namun terdapat kendala pada pembahasan hasil pemotongan menggunakan *google speech API* (*google recognizer*). Oleh karena itu, penulis akan melakukan penelitian lainnya yang berkaitan dengan pengenalan pemotongan suara. Penelitian selanjutnya yang akan dilakukan akan menggunakan teknik pemotongan kata selain *google recognizer* atau melakukan penelitian dengan studi kasus lain seperti menggunakan bahasa melayu Riau.

Daftar Pustaka

[1] A. S. M. BA WAZIR and J. H. CHUAH, "Spoken Arabic Digits Recognition Using Deep Learning," in *2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS 2019)*, Selangor, Malaysia, 2019.

[2] M. H. Tambunan, Martin, H. Fakhruroja, Riyanto and C. Machbub, "Indonesian Speech Recognition Grammar Using Kinect 2.0 for Controlling Humanoid Robot," in *The 2018 International Conference on Signals and Systems (ICSigSys)*, Bali, Indonesia, 2018.

[3] S. Duduka, H. Jain, V. Jain, H. Prabhu and P. P. M. Chawan, "Accent Classification using Machine Learning," in *International Research Journal of Engineering and Technology (IRJET)*, 2020.

[4] A. Ahmed, P. Tangri, A. Panda, D. Ramani and S. Karmakar, "VFNet: A Convolutional Architecture for Accent Classification," in *2019 IEEE 16th India Council International Conference (INDICON)*, 2019.

[5] L. M. A. Sheng and M. W. X. Edmund, "Deep Learning Approach to Accent Classification," in *stanford.edu*, 2017.

[6] K. Nugroho, "Javanese Gender Speech Recognition Based on Machine Learning Using Random Forest and Neural Network," *SISFORMA: Journal of Information Systems (e-Journal)*, vol. 6, 2019.

[7] J. A. Zea and J. Aguiar, "“Spanish Políglota”: an automatic Speech Recognition system based on HMM," in *2021 Second International Conference on Information Systems and*

Software Technologies (ICI2ST), Quito, Ecuador, 2021.

[8] J. Choi, H. Gill, S. Ou, Y. Song and J. Lee, "Design of voice to text conversion and management program based on Google Cloud Speech API," in *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, NV, USA, 2018.

[9] A. Winursito, R. Hidayat and A. Bejo, "Improvement of MFCC Feature Extraction Accuracy Using PCA in Indonesian Speech Recognition," in *2018 International Conference on Information and Communications Technology (ICOIACT)*, Yogyakarta, Indonesia, 2018.

[10] A. Winursito, R. Hidayat and A. Bejo, "Improvement of MFCC Feature Extraction Accuracy Using PCA in Indonesian Speech Recognition," in *2018 International Conference on Information and Communications Technology (ICOIACT)*, Yogyakarta, Indonesia, 2018.

[11] R. Hidayat, A. Bejo, S. Sumaryono and A. Winursito, "Denoising Speech for MFCC Feature Extraction Using Wavelet Transformation in Speech Recognition System," in *2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE)*, Bali, Indonesia, 2018.

[12] C. Jeyalakshmi, B. Murugeshwari and M. Karthick, "HMM and K-NN based Automatic Musical Instrument Recognition," in *2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)*, Palladam, India, 2018.

[13] N. Song and H. Yang, "A Gesture-to-Emotional Speech Conversion by Combining Gesture Recognition and Facial Expression Recognition," in *2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia)*, Beijing, China, 2018.

[14] W. Ting, "An Acoustic Recognition Model for English Speech Based on Improved HMM Algorithm," in *2019 11th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, Qiqihar, China, 2019.

[15] V. Gupta, R. S. Shankar and H. D. Kotha, "Voice Identification in Python Using Hidden Markov Model," *International Journal of Advanced Science and Technology*, vol. Vol. 29, 2020.

https://doi.org/10.31849/digitalzone.v13i2.10579