

Klasifikasi Gender dan Usia berdasarkan Suara Pembicara Menggunakan Hidden Markov Model

Irfan Tri Handoko ¹, Suyanto ²

School of Computing, Telkom University

Jl. Telekomunikasi Terusan Buah Batu, Bandung, West Java 40257, Indonesia

¹handokomon@student.telkomuniversity.ac.id, ²suyanto@telkomuniversity.ac.id

Abstract

Gender-age classification based on voice is useful in the introduction of speech such as emotional classification has better performance when separate acoustic models are used to determine male and female. Gender classification has also been implemented in face recognition, video summarization, determining different permission levels for different age groups, and others. Grouping different ages are divided into three groups: child, young, middle, and senior based on a certain age range. This study concentrates on the classification of gender and age groups based on the speaker's voice. The method applied is the Gaussian Mixture Model-Hidden Markov Model (GMM-HMM). First, build feature vectors using Mel-Frequency Cepstrum Coefficient (MFCC). Furthermore, conducting training to produce an acoustic model for all speakers in the training database including male and female speakers of various ages. Then, HMM was applied to detect gender and age groups from speaking data. The voice data comes from Common Voice sites containing user-sent blog posts, old books, films, and other public corporate speeches. The main objective is to enable training and testing of automatic speech recognition systems. This study shows the highest accuracy of 96.4% based on test data using GMM-HMM.

Keywords: Classification, Mel-Frequency Cepstrum Coefficient, Acoustic Models, Gaussian Mixture Model, Hidden Markov Model

Abstrak

Klasifikasi usia-jenis kelamin berdasarkan suara berguna dalam pengenalan pidato seperti klasifikasi emosional memiliki kinerja yang lebih baik ketika model akustik yang terpisah digunakan untuk menentukan pria dan wanita. Klasifikasi jenis kelamin juga telah diterapkan dalam pengenalan wajah, peringkasan video, menentukan tingkat izin yang berbeda untuk kelompok umur yang berbeda, dan lainnya. Pengelompokan usia yang berbeda dibagi menjadi tiga kelompok: anak, muda, menengah, dan senior berdasarkan rentang usia tertentu. Penelitian ini berkonsentrasi pada klasifikasi jenis kelamin dan kelompok umur berdasarkan suara pembicara. Metode yang digunakan adalah *Gaussian Mixture Model-Hidden Markov Model* (GMM-HMM). Pertama, membangun vektor ciri menggunakan *Mel-Frequency Cepstrum Coefficient* (MFCC). Selanjutnya, melakukan pelatihan untuk menghasilkan model akustik untuk semua penutur dalam database pelatihan termasuk penutur pria dan wanita dari berbagai usia. Kemudian, HMM diterapkan untuk mendeteksi jenis kelamin dan kelompok umur dari data pembicara. Data suara berasal dari situs Common Voice berisi posting blog yang dikirimkan pengguna, buku-buku lama, film, dan korpora pidato publik lainnya. Tujuan utamanya adalah untuk memungkinkan pelatihan dan pengujian sistem pengenalan suara otomatis. Studi ini menunjukkan akurasi tertinggi 96,4% berdasarkan data uji menggunakan GMM-HMM.

Kata Kunci: Klasifikasi, *Mel-Frequency Cepstrum Coefficient*, *Acoustic Models*, *Gaussian Mixture Model*, *Hidden Markov Model*

I. PENDAHULUAN

KLASIFIKASI jenis kelamin dan usia sudah sering digunakan berdasarkan objek gambar dengan hasil yang lebih matang, namun menggunakan objek berupa suara masih dalam tahap pengembangan.

Klasifikasi jenis kelamin dipecah menjadi dua yaitu suara pria dan wanita. Klasifikasi ini berguna sebagai dasar pengenalan karakteristik ucapan dan emosi ucapan pembicara. Beberapa penelitian mengungkap bahwa *pitch* seorang perempuan memiliki nilai lebih tinggi dibandingkan dengan seorang laki-laki yang dijadikan sebagai fitur untuk klasifikasi jenis kelamin. Perbedaan suara wanita dan pria secara matematis tidak jauh berbeda [18]. Namun, karena sejak kecil manusia sudah dibiasakan untuk membedakan mana suara seorang pria dan mana yang suara wanita, manusia menjadi sangat sensitif terhadap perbedaan suara wanita dan pria. Para ahli mengatakan frekuensi suara pria berada pada rentang 65 hingga 260 *Hertz*. Sementara frekuensi suara wanita tercatat pada rentang 100 sampai 525 *Hertz*. Ini berarti pria dan wanita dengan frekuensi suara 100 sampai 260 *Hertz* seharusnya sulit dibedakan kalau hanya didengar dari suaranya saja [18].

Klasifikasi suara seseorang berdasarkan usianya masih tergolong sulit [8] dan memiliki akurasi yang belum baik dikarenakan berdasarkan hasil analisis hubungan *pitch* terhadap usia menunjukkan bahwa terjadi penurunan nilai *pitch* rata-rata seiring bertambahnya usia. Faktor-faktor yang membatasi sistem pengenalan ucapan pembicara yaitu ketidakmampuan untuk mengidentifikasi fitur yang sensitif dan belum cukup kuat untuk mengakomodasi perbedaan artikulasi pembicara, variasi prosodi, dan perbedaan saluran suara [15].

Pada penelitian ini fokus pada pembangunan sistem klasifikasi suara dan jenis kelamin berdasarkan suara dengan membandingkan pengaruh jumlah koefisien frekuensi *cepstral* menggunakan MFCC yang dianggap sebagai ekstraktor fitur yang berhasil digunakan dalam pengenalan suara [11], kemudian *Gaussian Mixture Model-Hidden Markov Model* (GMM- HMM) digunakan untuk memecahkan permasalahan klasifikasi jenis kelamin dan usia berdasarkan suara dengan menentukan jumlah keadaan tersembunyi untuk diobservasi secara tepat. *Gaussian Mixture Model- Hidden Markov Model* (GMM-HMM) dianggap sebagai salah satu sistem *state-of-art* dan pengklasifikasi yang terbukti sangat efisien dalam menganalisis masalah dengan ruang fitur yang kompleks [10]. Sehingga dalam penelitian ini, metode yang diusulkan mampu membedakan suara pria dan wanita dari berbagai kelompok usia dengan menghasilkan akurasi yang lebih baik.

II. STUDI LITERATUR

Beberapa tahun terakhir ini, pemrosesan ucapan semakin penting, model pembelajaran mesin dan teknik penambangan data telah banyak digunakan untuk pengenalan jenis kelamin dan usia melalui suara. Model prediksi ini dapat mengidentifikasi jenis kelamin dan usia seseorang dengan memanfaatkan berbagai fitur seperti panjang pita suara, gaya berjalan, dan ucapan. Lebih khusus, sifat akustik yang diperoleh dari sinyal suara dan bicara seperti durasi, intensitas dan frekuensi dapat digunakan sebagai fitur untuk mengenali jenis kelamin dan usia pembicara. Sejumlah penelitian telah dilakukan dalam beberapa tahun terakhir; beberapa hasil yang bermanfaat disajikan secara singkat di bawah ini.

Penelitian pemrosesan ucapan mulai digunakan untuk identifikasi otomatis usia dan jenis kelamin seseorang dari pidatonya di forensik [13], biometrik suara [12], dan berbagai aplikasi lain dari interaksi manusia-komputer seperti perbaikan kepuasan pelanggan [2]. Untuk sinyal suara, ciri-ciri besaran biasanya merupakan keluaran dari beberapa bentuk teknik analisis spektrum seperti *Linear Predictive Coding* (LPC) atau *Mel-Frequency Cepstrum Coefficients* (MFCC) [5]. MFCC adalah salah satu fitur yang paling sering digunakan di bidang identifikasi atau verifikasi pembicara serta estimasi usia pembicara [7]. Filter ini digunakan untuk menangkap karakteristik fonetis penting dari sinyal ucapan [9].

Gaussian Mixture Model (GMM) dikenal sebagai salah satu metode yang dapat digunakan untuk klasifikasi otomatis usia dan jenis kelamin pembicara [6]. GMM menganalisis dan membandingkan pengaruh jumlah campuran yang berbeda dan berbagai jenis fitur bicara yang digunakan untuk klasifikasi jenis kelamin dan usia. Ketergantungan kompleksitas komputasi pada jumlah campuran yang digunakan juga dianalisis. Hasil evaluasi objektif dan subyektif menghasilkan rata-rata akurasi sebesar 95,2%. Kelemahannya adalah fungsi penentuan kelas usia beroperasi dengan model yang berbeda dilatih untuk suara pria dan perempuan. Dalam hal akurasi klasifikasi usia yang dihasilkan menurun dikarenakan pengaruh faktor-faktor perbedaan suara.

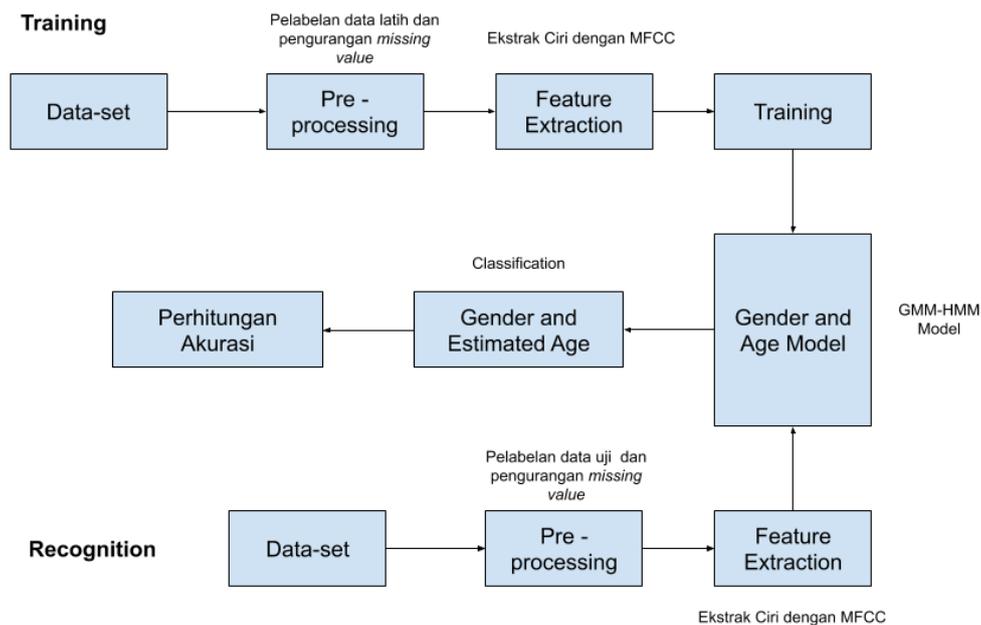
Berdasarkan riset yang dilakukan oleh Dr. Yusra Al-Irhayim dari *College of Computer Science and Mathematics, University of Mosul, Mosul Iraq* [3], klasifikasi jenis kelamin dan usia dapat dilakukan

menggunakan metode *Hidden Markov Model* (HMM). Konsep Hidden Markov memperluas model dengan memisahkan urutan pengamatan dan urutan keadaan. Untuk setiap keadaan, distribusi probabilitas didefinisikan yang menentukan seberapa besar kemungkinan setiap simbol pengamatan dihasilkan dalam keadaan tertentu. Tujuan dari penelitian tersebut adalah untuk meningkatkan tingkat pengakuan dengan mencari tahu parameter fitur yang baik berdasarkan MFCC dan klasifikasi menggunakan model *Hidden Markov*. Jadi, pada penelitian ini menguji pendekatan jenis kelamin-usia berdasarkan *Gaussian Mixture Model-Hidden Markov Model* (GMM-HMM) dengan ekstraktor ciri menggunakan MFCC diharapkan dapat meningkatkan keakuratan *state-of-the-art* dari metode pengenalan usia penutur dengan signifikansi statistik serta menganalisis pengaruh faktor-faktor utama yang mempengaruhi sistem estimasi jenis kelamin dan usia secara otomatis. Metode GMM-HMM dianggap sebagai model akustik yang sangat menjanjikan dengan mencapai hasil pengenalan suara yang baik [16]. Metode ini dapat mengenali karakteristik yang berbeda dari pembicara berdasarkan pola rekaman suara yang memfasilitasi untuk mengidentifikasi pembicara.

III. METODE PENELITIAN

A. Sistem yang dibangun

Sistem klasifikasi jenis kelamin dan usia berdasarkan suara yang dibangun menggunakan GMM-HMM terdiri dari dua langkah utama yaitu *training* dan *recognition*. Penelitian ini dapat diilustrasikan dengan mengikuti sebuah diagram alir sistem seperti ditunjukkan pada Gambar 1.



Gambar 1. Diagram Alir Sistem

Langkah pertama yang dilakukan dalam penelitian ini adalah pengumpulan data kemudian *pre-processing* data ucapan yang memiliki kategori jenis kelamin dan usia secara spesifik. Lalu meneliti karakteristik suara menggunakan MFCC untuk menghasilkan vektor ciri data pembicara. Setelah mendapat fitur ciri dari data ucapan kemudian memodelkan data dengan menghitung skor probabilitas tiap kategori menggunakan model campuran Gaussian dan dilatih menggunakan Hidden Markov Model serta melakukan perhitungan akurasi berdasarkan jumlah data yang benar dibagi jumlah seluruh data.

B. Pengumpulan Data

Pada penelitian ini digunakan basis data pidato utama Common Voice [14]. Common Voice adalah kumpulan data pidato yang dibaca oleh pengguna di situs web Common Voice berdasarkan pada teks dari sejumlah sumber domain publik seperti posting blog yang dikirimkan pengguna, buku-buku lama, film, dan korpora pidato publik lainnya. Tujuan utamanya adalah untuk memungkinkan pelatihan dan pengujian sistem pengenalan suara otomatis yang diucapkan oleh pembicara dalam keadaan netral. Sinyal audio stereo MP3 asli yang disampel pada 44,1 kHz dikonversi menjadi sinyal audio mono PCM 16-bit yang disetel ulang menjadi 22 kHz. Korpus yang digunakan penelitian ini berjumlah 9815 data ucapan pembicara dibagi menjadi dua bagian, yaitu 9063 untuk pelatihan dan 752 untuk pengujian.

C. Pra-Pemrosesan

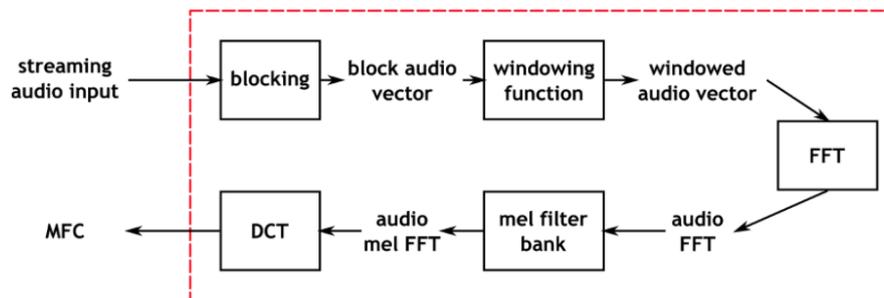
Data yang diperoleh kemudian diproses untuk menghilangkan data yang *missing value* seperti tidak tercantum label untuk usia atau jenis kelamin pembicara. Tujuannya adalah untuk mempermudah dalam melakukan pelabelan kelompok usia dan jenis kelamin serta mengoptimalkan pemodelan data. Berikut pelabelan untuk kelompok usia dari kedua jenis kelamin [6].

1. *Child – up to 12 years,*
2. *Young male/female – up to 25 years,*
3. *Middle male/female – up to 55 years,*
4. *Senior male/female – over 55 years.*

Data utama dengan kategori jenis kelamin dibagi menjadi 4 subkategori usia, yaitu *child, young, middle, senior*. Khusus kategori *child* tidak dibedakan kategorinya berdasarkan jenis kelamin sehingga terdapat 7 kategori yang akan diteliti.

D. Ekstraksi Fitur

Data yang sudah dibagi ke dalam beberapa kategori sesuai pelabelan kemudian diekstrak menggunakan MFCC. Proses tersebut digunakan untuk mendapatkan sederetan besaran sinyal input untuk menetapkan pola pembelajaran atau pola uji. Oleh karena itu, penyusun telah memilih untuk menggunakan MFCC untuk fitur penyusun. MFCC didasarkan atas variasi *bandwidth* kritis terhadap frekuensi pada telinga manusia. Filter ini bekerja secara linier pada frekuensi rendah dan logaritmik pada frekuensi tinggi [4] yang dianggap dapat menangkap karakteristik fonetis penting dari sinyal ucapan [9].



Gambar 2. MFCC (*Mel-Frequency Cepstrum Coefficients*) [9]

Bandwidth kritis adalah filter bandpass, disesuaikan di sekitar frekuensi pusat. Di bawah 1 kHz pita kritis ditempatkan linier di sekitar 100, 200, ... 1000 Hz. Di atas 1 kHz pita-pita ini ditempatkan dengan skala mel. Dalam perhitungan MFCC, energi total di setiap pita kritis digunakan, dengan menggunakan persamaan [17].

$$Y(i) = \sum_{k=0}^{N/2} \log |s(n)| \cdot H_i(k, \frac{2\pi}{N^i})$$

di mana $Y(i)$ adalah total energi dalam pita kritis, N adalah panjang gelombang, $S(n)$ adalah sinyal yang dihitung MFCC, $H_i()$ adalah filter pita kritis pada koefisien ke- i dan N adalah jumlah poin yang digunakan dalam jangka pendek.

Secara khusus, MFCC sangat populer dengan *Gaussian Mixture Model - Hidden Markov Models* (GMM-HMM) digunakan secara bersama-sama dan berkembang menjadi cara standar dalam melakukan *Automatic Speech Recognition* (ASR) [5]. Pengujian dilakukan dengan menentukan besar parameter MFCC untuk menghasilkan fitur suara yang tepat dalam klasifikasi seperti *winlen* (panjang dari analisis *window* dalam hitungan detik), *winstep* (langkah tiap *windowing* per detik), *nfilt* (jumlah *filter bank*), *numcep* (jumlah *cepstrum*) dan *nfft* yang merupakan ukuran FFT (*Fast Fourier Transform*). Penentuan *numcep* dan *nfilt* saling terkait di mana *nfilt* yang digunakan merupakan hasil kali dua dengan *numcep*. FFT digunakan untuk meminimalkan diskontinuitas pada sinyal yang dipengaruhi oleh besar frekuensi tiap sampel suara dan *winlen*.

E. Gaussian Mixture Model

Vektor ciri yang sudah diperoleh berdasarkan MFCC kemudian dimodelkan dengan *Gaussian Mixture Model* (GMM). Model campuran *Gaussian* dapat didefinisikan sebagai kombinasi linier dari beberapa fungsi distribusi probabilitas *Gaussian* dari vektor data input. Dalam pembuatan GMM, perlu untuk menentukan matriks kovarians, vektor nilai rata-rata, dan parameter bobot dari data pelatihan input. Fungsi kemungkinan maksimum GMM ditemukan menggunakan algoritma iterasi ekspektasi-maksimalisasi [13] yang dikendalikan oleh jumlah campuran dan jumlah langkah iterasi. Sebagian besar pengklasifikasi GMM menggunakan skor akhir dari model yang diberikan oleh probabilitas keseluruhan maksimum untuk kelas yang sesuai [3].

$$m^* = \arg \max_{1 \leq n \leq N} \text{score}$$

Di mana skor (T, n) mewakili nilai probabilitas pengelompokan GMM untuk model yang dilatih untuk kelas ke- n saat ini dalam proses evaluasi, N adalah jumlah kelas, dan T adalah vektor input dari fitur yang diperoleh dari data ucapan yang diuji. Pendekatan yang relatif sederhana dan kuat ini tidak dapat mencapai akurasi pengakuan terbaik dalam semua kasus. Oleh karena itu, metode ini diperpanjang dengan perhitungan skor akumulasi untuk menentukan keputusan tentang jenis kelamin-usia pembicara yang dinyatakan sebagai

$$m_{acc} = \arg \max_{1 \leq i \leq M} \prod_{p=1}^P (m^*(i, p) \equiv 1)$$

F. Hidden Markov Model

Selain melakukan perhitungan skor probabilitas menggunakan GMM, *Hidden Markov Model* (HMM) digunakan untuk memperluas model dengan memisahkan urutan observasi dan urutan keadaan. Asal mula HMM adalah rantai teori probabilitas *Markov* yang terkenal yang dapat digunakan untuk pemodelan berurutan. HMM adalah *finite state automaton* yang memiliki sejumlah keadaan yang terbatas di mana mesin membuat transisi dari satu keadaan ke keadaan lain. Model ini juga memiliki *set input* dan *output* yang terbatas.

Ketika sebuah mesin berada pada keadaan saat itu tergantung pada simbol input yang diberikan, mesin akan berpindah ke keadaan lain yang memancarkan *output* tertentu dengan probabilitas tertentu pada saat berikutnya secara instan. HMM banyak digunakan dalam pengenalan suara karena kemampuannya untuk merepresentasikan aspek waktu dari sinyal ucapan [1]. Masing-masing keadaan distribusi probabilitas

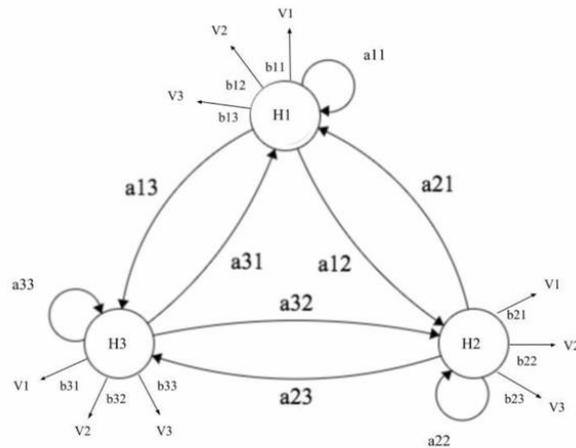
ditentukan oleh seberapa besar kemungkinan setiap simbol pengamatan akan dihasilkan dalam keadaan tertentu. Hidden Markov Model dapat didefinisikan oleh parameter berikut [16].

H : Set of hidden states

V : Set of visible states

a_{ij} : Transition probabilities corresponding to the visible states.

b_{jk} : Emission probability of the visible state from the hidden states



Gambar 3. HMM Model [16]

States biasanya akan diindikasikan oleh i, j menyatakan bahwa 'the model is in ' pada suatu titik waktu tertentu t yang akan ditunjukkan oleh q_t . Jadi, $q_t = i$ menyatakan bahwa model dalam keadaan i pada waktu t . Distribusi probabilitas transisi antar states $\{ij\}$. $A = a$, di mana

$$a_{ij} = P(q_{i+1} = j | q_i = i) \quad 1 \leq i, j \leq N$$

HMM dalam penelitian ini digunakan untuk menghasilkan model *training* klasifikasi jenis kelamin-usia di mana model ini melakukan sejumlah iterasi sebanyak 100 kali untuk melakukan observasi antar urutan keadaan yang diamati. Model yang dibuat menggunakan matriks kovarians secara penuh dengan meningkatkan jumlah *states* yang akan diteliti. Model ini dapat melakukan pengambilan sampel dan evaluasi dengan mudah serta estimasi kemungkinan maksimum dari parameter.

G. Evaluasi

Penelitian ini dilakukan dalam 4 kali percobaan terbaik secara berbeda. Dua percobaan pertama digunakan untuk menentukan variabel MFCC yang berpengaruh dalam klasifikasi (*winlen*, *winstep*, *numcep*, *nfilt*, dan *nfft*) dengan menggunakan parameter *default* GMM-HMM. Setelah menentukan nilai fitur cepstral yang berpengaruh dalam pengenalan suara, dua percobaan berikutnya dengan melakukan pengujian terhadap parameter GMM-HMM seperti jumlah keadaan, iterasi, tipe, dan bobot kovarians serta menentukan algoritma yang tepat untuk mengurutkan keadaan tersembunyi dari HMM. Hasil performa dari sistem yang dibuat, dievaluasi berdasarkan akurasi dengan rumus.

$$Recognition_{acc} = \frac{N_{rec}}{N_{all}} \times 100\%$$

N_{rec} adalah jumlah data ucapan yang terklasifikasi dengan benar dan N_{all} adalah jumlah data ucapan keseluruhan yang diuji untuk klasifikasi.

IV. HASIL DAN DISKUSI

Setelah melakukan *training* pada data untuk menghasilkan model pelatihan, kemudian dilakukan proses pengujian terhadap data *test* yang sudah dilakukan *pre-processing* dan dikategorikan berdasarkan kelasnya. Untuk mengevaluasi performa yang diperoleh dengan melakukan perbandingan antara model yang diusulkan dengan model yang lain untuk mengklasifikasikan jenis kelamin dan usia berdasarkan suara pembicara.

TABEL I
HASIL EVALUASI KLASIFIKASI JENIS KELAMIN DAN USIA DENGAN GMM

Percobaan ke-n	1	2	3	4
<i>Child</i>	90.40 %	94.92%	94.90%	93.81
<i>Male</i> (Y + M + S)	96.86%	99.18%	99.2%	97.53%
<i>Female</i> (Y + M + S)	94.29%	98.15%	98.07%	96.39%
Akurasi (std)	94.8%	95.4%	95.4%	95.4%

TABEL II
HASIL EVALUASI KLASIFIKASI JENIS KELAMIN DAN USIA DENGAN GMM-HMM

Percobaan ke-n	1	2	3	4
<i>Child</i>	90.3%	93.2%	92.7%	95%
<i>Male</i> (Y + M + S)	93.2%	94.8%	96%	96.8%
<i>Female</i> (Y + M + S)	96%	94.2%	98.2%	97.4%
Akurasi (std)	93.2%	94.1%	95.6%	96.4%

Berdasarkan TABEL I dan TABEL II, klasifikasi model campuran *Gaussian* dengan penambahan model HMM sebagai klasifikasi dan ekstrak ciri sinyal audio menggunakan MFCC mencapai rata-rata akurasi maksimum sebesar 96.4% di mana lebih baik daripada model GMM biasa [6]. Intuisi tentang fitur cepstral dapat membantu mencari tahu apa yang harus dicari ketika menggunakannya dalam sistem berbasis ucapan. Jendela analisis diuji dengan menggunakan nilai *default* dari MFCC sebesar 0.025 yang digunakan sebagai penyesuaian ukuran FFT berdasarkan frekuensi tiap sampel suara. Langkah analisis tiap detiknya sebesar 0.05 *step* antar jendela berturut-turut yang mampu mencari karakteristik pengenalan suara disertai penentuan jumlah *filter bank* sebesar 26 *filter* yang dipengaruhi oleh nilai *cepstrum* yang diteliti.

Selain fitur cepstral, HMM memiliki peran dalam pengenalan suara secara akurat. Dengan menambah jumlah *states* di dalam model HMM memungkinkan untuk mengoptimalkan sistem pengenalan dan meningkatkan akurasi klasifikasi suara namun harus menentukan algoritma yang tepat untuk mengurutkan keadaan di dalam model karena semakin bertambahnya jumlah *states* yang diteliti, semakin kompleks pemodelannya. Dalam penelitian ini, model HMM dengan emisi *Gaussian* diuji dengan menggunakan bobot kovarians sebesar 1 dan tipe kovarians penuh agar setiap bagian keadaan memiliki matriks kovarians umum masing-masing. Rentang jumlah *states* yang digunakan di antara 7 sampai dengan 42 keadaan diobservasi dengan melakukan 100 kali iterasi. Dalam menangani masalah mendasar HMM, penelitian ini menggunakan algoritma Viterbi yang mampu mengingat parameter model dan data yang diamati, perkiraan kemungkinan urutan optimal dari keadaan tersembunyi, dan menghitung kemungkinan model.

V. KESIMPULAN

Penelitian ini mengembangkan sistem klasifikasi jenis kelamin dan usia berdasarkan data suara pembicara yang berasal dari situs Common Voice menggunakan GMM-HMM telah sukses diimplementasikan. Pemilihan fitur adalah salah satu faktor paling penting dalam merancang sistem pengenalan jenis kelamin dan usia serta MFCC telah dipilih sebagai fitur untuk pengenalan suara yang tepat. Pengujian menggunakan GMM-HMM dianggap sebagai salah satu sistem *state-of-art* dan pengklasifikasi yang terbukti sangat efisien dalam menangani masalah dengan ruang fitur yang kompleks sehingga bisa menghasilkan akurasi sebesar 96,4%. Namun, model ini masih dapat ditingkatkan lebih lanjut dengan penentuan parameter yang lebih tepat dan penggunaan *dataset* yang lebih besar. Manfaat dari model HMM ini adalah mengoptimalkan urutan keadaan dari faktor-faktor utama yang mempengaruhi sistem klasifikasi jenis kelamin dari berbagai kelompok usia secara otomatis.

PERNYATAAN

Kami mengucapkan terimakasih kepada orang tua, seluruh kolega di Universitas Telkom, khususnya di Fakultas Informatika, atas segala dukungannya dalam penyelesaian penelitian ini.

REFERENSI

- [1] D. B. D. B. Arafat Abu Mallouh. SA Framework for Enhancing Speaker Age and Gender Classification by Using a New Feature Set and Deep Neural Network Architectures. THE SCHOOL OF ENGINEERING UNIVERSITY OF BRIDGEPORT CONNECTICUT, 2017.
- [2] B. Barkana and J. Zhou. A New Pitch-Range Based Feature Set for a Speaker's Age and Gender Classification. Applied Acoustics, 98, 2015.
- [3] A. a. Dr. Yusra Al-Irhayim. Speaker Gender Recognition Using Hidden Markov Model. College of Computer Science and Mathematics, University of Mosul, Mosul Iraq, 2016.
- [4] F. Faek. Objective Gender and Age Recognition from Speech Sentences. Aro, The Scientific Journal of Koya University, 3 (2), 24-29. doi:10.14500/aro.10072, 2015.
- [5] H. Fayek. Speech Processing for Machine Learning: Filter banks, Mel-Frequency Cepstral Coefficients (MFCCs) and What's in-between. <https://haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html>. Online; Accessed 2 November 2018.
- [6] J. M. Jiri Pribil, Anna Pribilova. GMM-Based Speaker Age And Gender Classification In Czech And Slovak. Journal of ELECTRICAL ENGINEERING, 68:3-12, 2017.
- [7] J. M. Jiri Pribil, Anna Pribilova. GMM-Based Speaker Gender And Age Classification After Voice Conversion. Journal of ELECTRICAL ENGINEERING, 2017.
- [8] D. Katerenchuk. Age Group Classification With Speech And Metadata Multimodality Fusion. CUNY Graduate Center 365 Fifth Avenue, Room 4319 New York, USA, 2018.
- [9] D.Z.J.Z. W.Z.LianzhangZhu, LeimingChen. Emotion Recognition from Chinese Speech for Smart Affective Services Using a Combination of Svm and Dbn. Sensors 2017, 17, 1694; doi:10.3390/s17071694, 7, 2017.
- [10] Z. R. T. D. T. Lozano-Diez, A. and J. Gonzalez-Rodriguez. An Analysis of The Influence of Deep Neural Network (DNN) Topology in Bottleneck Feature Based Language Recognition. Plos One, 12(8). doi:10.1371/journal.pone.0182580, 2017.
- [11] J. Lyons. Mel Frequency Cepstral Coefficient (MFCC) Tutorial. <http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/>. Online; Accessed 15 Desember 2018.
- [12] M.E.M.Fairhurst and M.D.Costa-Abreu. Selective Review and Analysis of Aging Effects in Biometric System Implementation. IEEE Transactions on Human-Machine Systems, 45, 2015.
- [13] H.v.H, M.H.Bahari, M.McLarenand, D. Van Leeuwen. Speaker Age Estimation Using I-Vectors, Engineering Applications of Artificial Intelligence. Sensors 2017, 17, 1694; doi:10.3390/s17071694, 34, 2014.
- [14] K. D. Michael Henretty, Tilman Kamp, and T. C. V. Team. 500 Hours of Speech Recordings, with Speaker Demographics. <https://www.kaggle.com/mozillaorg/common-voice>. Online; Accessed 3 November 2018.
- [15] T. S. K. D. K. A. Z. I. S. C. Nagendra Kumar Goel, Mousmita Sarma. Extracting Speaker's Gender, Accent, Age and Emotional State from Speech. Go-Vivace Inc., McLean, VA, USA, 2018.
- [16] P. Saikia. Hmm-dnn speech recognition techniques: a review. Gauhati University-Institute of Distance and Open Learning, Assam, India, 7:14068-14072, 2017.
- [17] F. B. Vej. The Mel Frequency Scale and Coefficients. http://kom.aau.dk/group/04gr742/pdf/MFCC_worksheet.pdf. Online; Accessed 24 January 2019.
- [18] Voice-Academy. Male and Female Voices. <https://uiowa.edu/voice-academy/male-female-voices>. Online; Accessed 1 September 2018.