

PENDETEKSIAN TINGKAT USIA MUDA, DEWASA DAN TUA MENGGUNAKAN METODE MFCC DAN FUZZY LOGIC BERBASISKAN SPEECH RECOGNITION

DETECTION OF THE LEVEL OF YOUTH, ADULTS AND ELDERLY BY USING MFCC METHOD AND FUZZY LOGIC BASE ON SPEECH RECOGNITION

Restu Wardani¹, Dr.Ir. Bambang Hidayat, DEA², Suci Aulia, S.T., M.T.,³

¹Prodi S1 Teknik Telekomunikasi, Fakultas Teknik Elektro, Universitas Telkom

²Prodi D3 Teknik Telekomunikasi, Fakultas Ilmu Terapan, Universitas Telkom

¹restuwardani@student.telkomuniversity.ac.id ²bhidayat@telkomuniversity.ac.id

³suciaulia@telkomuniversity.ac.id

Abstrak

Melalui suara seseorang akan mendapatkan informasi yang dibutuhkan dan mengetahui siapa yang sedang berbicara. Namun dewasa ini, suara bukanlah menjadi suatu parameter faktor keamanan dalam hal berkomunikasi dikarenakan banyaknya terjadi penipuan melalui telepon yang hanya memanfaatkan suara penelepon tanpa tahu identitasnya. Pendeteksian suara adalah salah satu metode keamanan berkomunikasi dengan membedakan suara manusia yang berusia muda, dewasa dan tua. Dengan pendeteksian suara, dapat mempersempit identifikasi seseorang. Dalam tugas akhir ini, tahap yang dilakukan adalah akuisisi data, *pre-processing*, ekstraksi ciri dan klasifikasi dengan masukan sinyal berupa sinyal bicara. Untuk ekstraksi ciri, metode yang digunakan adalah *Mel-Frequency Cepstral Coefficient* (MFCC). Setelah didapatkan ciri dari sinyal bicara tersebut, selanjutnya akan dilakukan metode klasifikasi untuk mencocokkan dengan ciri dari sinyal bicara yang telah didapatkan dengan menggunakan *Fuzzy Logic*. Jumlah data latih yang digunakan dalam Tugas Akhir ini adalah sebanyak 84 dan 84 data untuk data uji. Hasil dari penelitian Tugas Akhir ini adalah didapatkan nilai akurasi sebesar 94.05% dengan waktu komputasi selama 1.59 detik. Sedangkan untuk waktu komputasi tercepat yaitu selama 0.49 detik dengan nilai akurasi sebesar 76.19%.

Kata Kunci : *speech recognition, mel-frequency cepstral coefficient, fuzzy logic*

Abstract

Through the voice of someone will get the needed information and knowing who is talking. But today, the sound is not be a parameter in terms of communicating the safety factor because the number of fraud over the phone which utilizes only the caller's voice without knowing his identity. Sound detection is one of the security methods distinguish the human voice to communicate with the young, adults and the elderly. With sound detection, can narrow the identification of a person. In this final project, the phases are data acquisition, *pre-processing*, feature extraction and classification of the input signal in the form of speech signals. For feature extraction, the method used is the *Mel-Frequency cepstral coefficient* (MFCC). Having obtained the characteristics of the speech signal, the classification method will be done to match the characteristics of the speech signal which has been obtained by using *Fuzzy Logic*. The amount of training data used in this final project is as much as 84 and 84 data for testing data. Final results of this final project are the values obtained with an accuracy of 94.05% for 1.59 seconds of computing time. As for the fastest computing time is for 0.49 seconds with an accuracy of 76.19% value.

Keywords: *speech recognition, mel-frequency cepstral coefficient, fuzzy logic*

1. PENDAHULUAN

Suara menjadi hal penting dikarenakan melalui suara seseorang dapat mengetahui siapa yang sedang berbicara dan informasi apa yang telah didapatkan. Di era globalisasi seperti saat ini, banyak teknologi yang memanfaatkan suara manusia dalam pengaplikasiannya. Namun terkadang teknologi yang ada belum dapat membantu masyarakat dalam menjalin komunikasi. Seperti dalam hal berkomunikasi melalui telepon, terkadang masyarakat masih memiliki kebingungan dengan membedakan suara orang diseborang telepon yang sedang berbicara. Perbedaan suara ini akan sangat membantu jikalau terjadi penipuan lewat telepon. Melalui suara, dapat dikenali bahwa suara orang yang sedang berbicara tersebut apakah suara orang muda, tua atau dewasa. Dengan mengenali suaranya, maka akan dengan mudah mempersempit identifikasi orang tersebut. Proses

identifikasi dengan cara mendeteksi suara dapat dilakukan dengan cara membedakan tipe suara antara pria dan wanita [1]. Pendeteksian yang dilakukan pada tugas akhir ini adalah dengan cara membedakan suara antara suara manusia yang berada di tingkat usia muda, dewasa atau tua. Banyak metode yang digunakan dalam pendeteksian suara seperti mengkombinasikan metode *Mel Frequency Cepstral Coefficient* (MFCC) sebagai ekstraksi cirinya dan *Support Vector Machine* (SVM) sebagai klasifikasinya[4]. Tugas akhir ini hanya akan menggunakan metode *Mel Frequency Cepstral Coefficient* (MFCC) dan *Fuzzy Logic* sebagai klasifikasinya untuk mendeteksi tingkat usia muda, dewasa dan tua. Tingkat usia yang dimaksudkan dalam tulisan ini adalah tingkat usia orang tua, dewasa dan muda. Usia muda yaitu usia sekitar 17-30 tahun, usia dewasa yaitu usia sekitar 31-50 tahun, sedangkan usia tua yaitu dimulai dari usia 51 tahun keatas[2]. Dengan mendeteksi tingkat usia muda, dewasa dan tua hanya melalui suara maka akan mempersempit penyelidikan terhadap identitas seseorang.

2. DASAR TEORI

A. Audio

Audio atau suara adalah getaran udara pada frekuensi yang dapat didengar oleh telinga manusia sehingga disebut dengan frekuensi suara atau frekuensi audio. Frekuensi audio berada diantara 20 Hz sd 20 KHz. Karakteristik suara ditentukan antara lain oleh frekuensi, amplitudo dan durasi. Ada dua jenis audio yaitu audio analog dan audio digital. Audio analog adalah pengolahan suara asli (akustik) melalui peralatan elektronik analog sedangkan audio digital adalah suara yang melalui pengolahan secara digital melalui komputer[6].

B. Konsep Dasar *Speech Recognition*

Pengenalan suara atau *speech recognition* adalah proses konversi sebuah sinyal akustik, yang berasal dari *microphone* menjadi satu atau sekumpulan kata. Pengenalan suara merupakan upaya agar manusia dan mesin dapat berkomunikasi melalui media suara. Tujuan dari teknologi pengenalan suara adalah bagaimana menciptakan suatu mesin yang dapat menerima dan mengerti informasi yang terdapat pada sinyal bicara dan melakukan suatu tindakan sesuai informasi yang terkandung pada sinyal suara tersebut. Secara umum, pada *speech recognizer* akan memproses sinyal suara yang masuk dan menyimpannya dalam bentuk digital. Hasil proses digitalisasi tersebut kemudian dikonversi dalam bentuk spektrum suara yang akan dianalisa dengan membandingkannya dengan template suara pada *database* sistem[7].

C. Mel-Frequency Cepstral Coefficient

Mel frequency Cepstral Coefficient (MFCC) merupakan salah satu metode ekstraksi ciri sinyal suara yang berdasarkan prinsip karakteristik pendengaran telinga manusia. Kemampuan pendengaran pada telinga manusia tidak berskala linier namun dihitung dalam skala ‘*Mel*’ yang disebut sebagai ‘*Mel Scale*’. Skala *mel-frequency* adalah frekuensi linear berada dibawah 1000 Hz dan bentuk logaritmik berada diatas 1000 Hz. Sebagai titik referensi adalah *pitch* dengan *tone* 1 kHz, 40 dB diatas nilai batas ambang pendengaran, ini dinyatakan 1000 *mel*. Pendekatan persamaan untuk menghitung *mel* dalam frekuensi f(Hz) dapat dilihat pada persamaan berikut.[7]

$$f_{mel} = 2592 * \log_{10} \left(1 + \frac{f}{700} \right) \dots\dots\dots(1)$$

$$M^{-1}(m) = 700 \left(\exp \left(\frac{m}{1125} \right) - 1 \right) \dots\dots\dots(2)$$

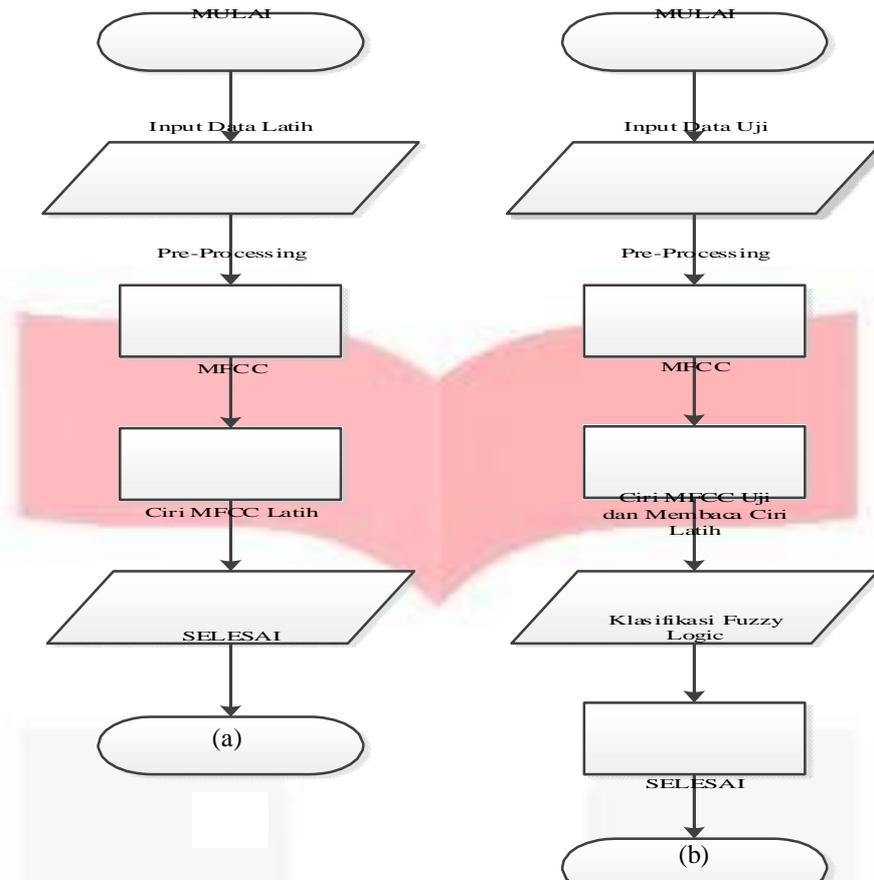
D. Fuzzy Logic

Logika *Fuzzy* adalah peningkatan dari logika *boolean* yang berhadapan dengan konsep kebenaran sebagian. Saat logika klasik menyatakan bahwa segala hal dapat diekspresikan dalam istilah biner (0 atau 1, hitam atau putih, ya atau tidak), logika *fuzzy* menggantikan kebenaran *boolean* dengan tingkat kebenaran. Logika *fuzzy* memungkinkan nilai keanggotaan antara 0 dan 1, tingkat keabuan dan juga hitam dan putih, dan dalam bentuk linguistik, konsep tidak pasti seperti “sedikit”, “lumayan”, dan “sangat”. Logika ini berhubungan dengan set *fuzzy* dan teori kemungkinan. Logika *fuzzy* diperkenalkan oleh Dr. Lotfi Zadeh dari Universitas California, Berkeley pada tahun 1965. Ada tiga proses utama jika ingin mengimplementasikan *fuzzy logic* pada suatu perangkat, yaitu fuzzifikasi, evaluasi *rule*, dan defuzzifikasi[3][8].

3. MODEL SISTEM DAN METODOLOGI PERANCANGAN

Dalam perancangan sistem pendeteksian ini sinyal yang diinputkan berupa sinyal bicara atau suara dengan durasi 2-3 detik. Suara pada data latih sistem merupakan suara dari 14 orang *speaker* yang terdiri dari 4 orang (satu orang laki-laki dan tiga orang perempuan) yang berada di tingkat usia muda, 5 orang (dua orang laki-

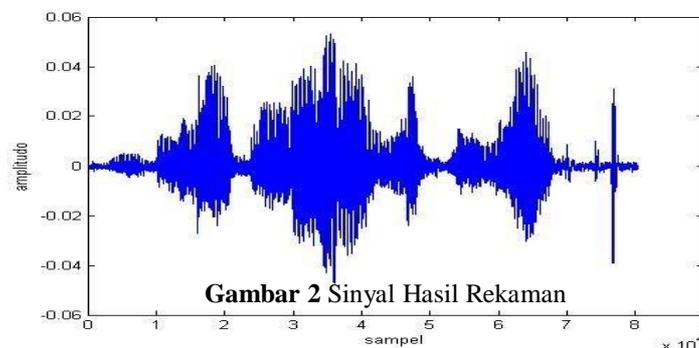
laki dan tiga orang perempuan) yang berada di tingkat usia dewasa, dan 5 orang (dua orang laki-laki dan tiga orang perempuan) yang berada di tingkat usia tua.



Gambar 1 Diagram Alir Perancangan Sistem. (a) Proses Latih (b) Proses Uji

A. Akuisisi Data

Akuisisi data merupakan tahap perekaman suara. Setiap *speaker* mengucapkan tiga kalimat. Masing-masing kalimat diucapkan sebanyak tiga kali. Frekuensi sampling yang digunakan adalah 44100 Hz. Format data suara disimpan dalam format *Waveform (*.wav)*.



Gambar 2 Sinyal Hasil Rekaman

B. Pre-Processing

- Inisialisasi daerah *silent* merupakan proses melakukan pencarian daerah *silent* berdasarkan nilai standar deviasi dari masing-masing sinyal suara. Elemen matriks yang bernilai kurang dari nilai standar deviasi

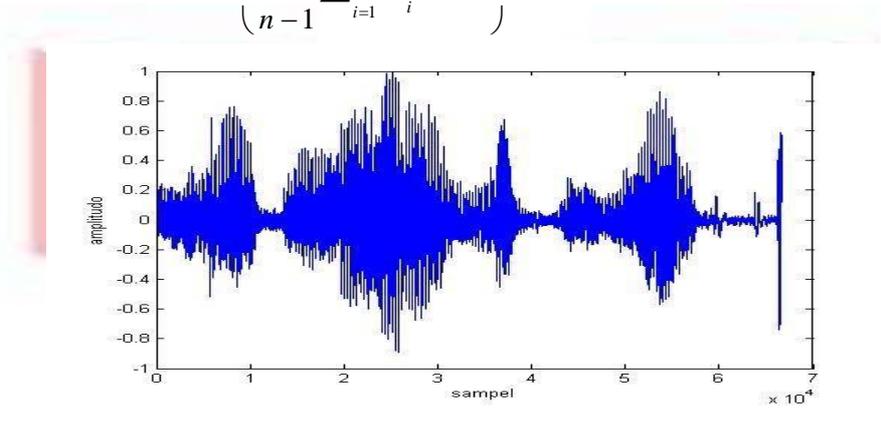
diinisialisasi sebagai daerah *silent* dimana daerah *silent* tersebut kemudian menjadi parameter *input noise*.

- Proses selanjutnya adalah normalisasi dimana proses ini merupakan proses untuk menyamakan amplituda dari setiap suara yang direkam oleh sistem sehingga berada dalam rentang -1 sampai +1.

$$X_{norm} = \frac{x}{\max(|x(n)|)} \dots\dots\dots (3)$$

- Selanjutnya adalah proses *Silence removal* yaitu proses untuk mencari dan menghilangkan daerah *silent* berdasarkan nilai standar deviasi dari masing-masing sinyal suara.[7]

$$std = \left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2} \dots\dots\dots (4)$$

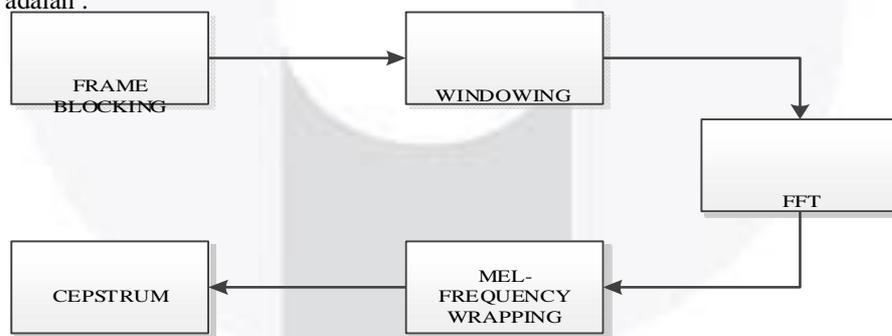


Gambar 3 Sinyal Hasil Normalisasi dan *Silence Removal*

- Data sinyal suara kemudian di filter menggunakan filter *pre-emphasis* untuk menyaring frekuensi tinggi serta meningkatkan *signal to noise ratio* dan menghilangkan *noise*.

C. Ekstraksi Ciri Dengan Mel-Frequency Cepstral Coefficient

Setelah sinyal diproses melalui tahap *pre-processing*, kemudian sinyal akan diproses ke tahap utama yaitu ekstraksi ciri untuk mendapatkan nilai ciri dari masing-masing sinyal data. Tahapan proses untuk mendapatkan nilai ciri tersebut adalah :



Gambar 4 Diagram Blok Proses MFCC

a. **Frame Blocking**

Pada proses ini sinyal dibagi ke dalam beberapa *frame* yang diinginkan yaitu 128, 256 dan 512 sampel. Proses *overlapping* antar *frame* juga dilakukan dimana panjang *overlapnya* adalah setengah dari panjang *frame*.

b. **Windowing**

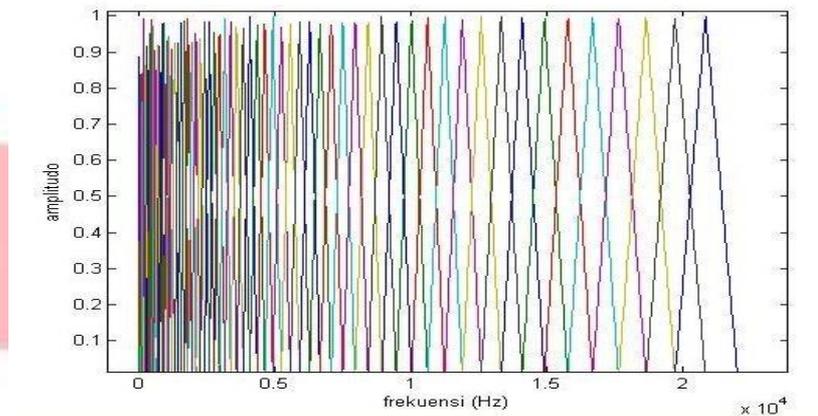
Proses *windowing* digunakan untuk mengurangi kebocoran *spectral* dan mengurangi efek diskontinuitas di awal dan akhir masing-masing *frame* yang dapat terjadi akibat proses *frame blocking*. *Window* yang digunakan dalam penelitian ini adalah *window hamming*.

c. **Fast Fourier Transform**

Penelitian ini menggunakan FFT sebagai pengkonversi dari domain waktu ke domain frekuensi. Jumlah titik FFT yaitu sebesar nilai kelipatan dua terdekat dengan jumlah sampel suatu *frame*.

d. **Mel-Frequency Wrapping**

Salah satu pendekatan simulasi spektrum yaitu menggunakan *filterbank*, satu *filter* untuk masing-masing komponen *mel-frequency* yang diinginkan. *Filterbank* mempunyai respon frekuensi *bandpass* segitiga dan jarak *bandwidth* ditentukan oleh konstanta interval *mel-frequency*. Proses untuk mendapatkan *filterbank* dimulai dari menentukan nilai frekuensi rendah dan frekuensi tinggi. Penelitian ini menggunakan frekuensi sampling 44100 Hz dengan nilai frekuensi rendah 0 Hz dan frekuensi tinggi 22050 Hz. Nilai-nilai skala Hz tersebut kemudian diubah ke dalam skala 'Mel', lalu rentang nilai tersebut dibagi menjadi beberapa nilai sejumlah *filterbank*, dalam hal ini *filterbank* yang digunakan adalah 64 buah. Rentang nilai dalam skala 'Mel' adalah 0 sampai 3906,19. Selanjutnya nilai-nilai hasil konversi 'Mel' tersebut di konversi kembali lagi ke dalam skala Hz.

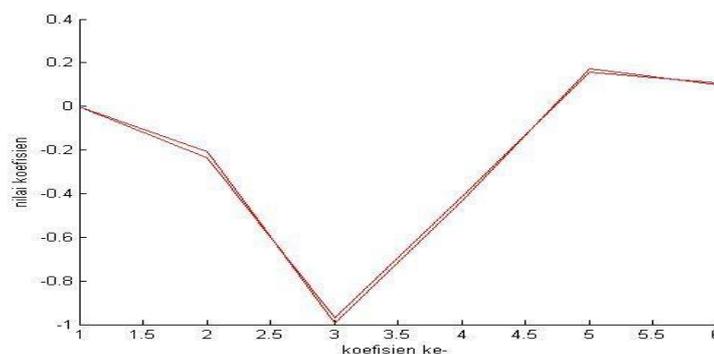


Gambar 5 Triangular Filterbank

Proses perkalian antara *Mel filterbank* dengan spektral daya periodogram hasil proses FFT menghasilkan energi dari *filterbank*. Hasil perkalian spektral daya dengan *filterbank* tersebut kemudian dihitung nilai *log* dari masing-masing energi. Setelah didapatkan hasilnya, kemudian dilakukan perhitungan DCT (*Discrete Cosine Transform*) untuk masing-masing energi tersebut.

e. **Cepstral Lifting**

Proses DCT akan menghasilkan koefisien MFCC dimana koefisien MFCC yang dihasilkan akan diperhalus melalui proses *cepstral liftering* sehingga akan lebih baik digunakan saat klasifikasi nanti.



Gambar 6 Hasil Ciri Tingkat Usia Tua

D. Klasifikasi Fuzzy Logic

Sistem *fuzzy* yang digunakan pada tahap klasifikasi ini adalah sistem ANFIS (*Adaptive Neuro-Fuzzy Inference System*) yang berupa jaringan adaptif berbasis sistem kesimpulan *fuzzy* (*fuzzy inference system*). Dengan penggunaan suatu prosedur *hybrid learning*, ANFIS dapat membangun suatu mapping *input-output* yang keduanya berdasarkan pada pengetahuan manusia (pada bentuk aturan *fuzzy if-then*) dengan fungsi keanggotaan yang tepat[5]. Tahap klasifikasi terdiri dari dua proses. Proses pertama yaitu proses latihan (*training*) dimana parameter dalam proses ini adalah *epoch* dan proses yang kedua adalah proses uji (*testing*) yaitu proses pengujian terhadap 84 data latihan dan 84 data uji.

4. HASIL DAN ANALISIS

Dilakukan beberapa skenario dalam pengujian tugas akhir ini, skenario-skenario tersebut adalah :

- **Skenario 1** : Pengujian akurasi dan waktu komputasi sistem berdasarkan lebar *frame* dan koefisien MFCC dengan *window hamming* dan nilai *epoch* sebanyak 10 kali

Tabel 1 Tabel Pengujian Lebar *Frame* dan Koefisien MFCC

nF	Koefisien					
	4		6		8	
	Akurasi	Waktu Komputasi	Akurasi	Waktu Komputasi	Akurasi	Waktu Komputasi
128	79.76 %	1.47 s	94.05 %	1.59 s	89.29 %	1.61 s
256	78.57 %	0.81 s	89.29 %	0.83 s	84.52 %	0.86 s
512	76.19 %	0.49 s	88.10 %	0.51 s	86.90 %	0.51 s

- **Skenario 2** : Pengujian akurasi dan waktu komputasi sistem berdasarkan lebar *frame* dan *epoch* dengan *window hamming* dan koefisien MFCC sebanyak 6 buah koefisien

Tabel 2 Tabel Pengujian Lebar *Frame* dan *Epoch*

nF	Epoch					
	20		100		150	
	Akurasi	Waktu Komputasi	Akurasi	Waktu Komputasi	Akurasi	Waktu Komputasi
128	92.86%	1.50 s	85.71%	1.54 s	85.71%	1.54 s
256	90.48%	0.85 s	89.29%	0.82 s	86.90%	0.82 s
512	91.67%	0.51 s	91.67%	0.51 s	91.67%	0.51 s

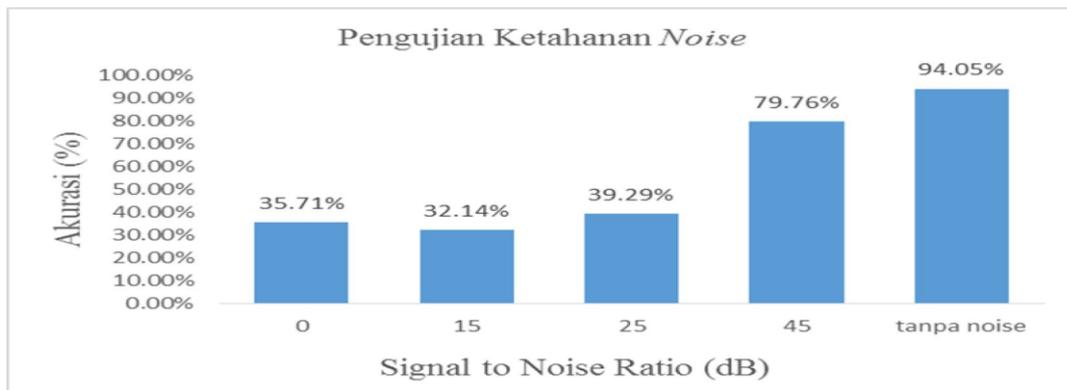
- **Skenario 3** : Pengujian akurasi dan waktu komputasi sistem berdasarkan koefisien MFCC dan *epoch* dengan *window hamming* dan lebar *frame* 128 sampel

Tabel 3 Tabel Pengujian Koefisien MFCC dan *Epoch*

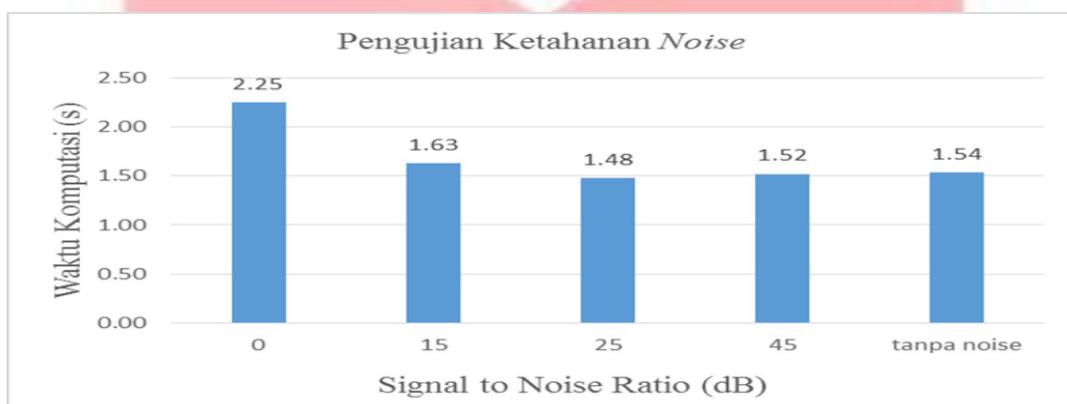
Coef	Epoch					
	20		100		150	
	Akurasi	Waktu Komputasi	Akurasi	Waktu Komputasi	Akurasi	Waktu Komputasi
4	80.95%	1.45 s	78.57%	1.45 s	79.76%	1.45 s
6	92.86%	1.50 s	85.71%	1.54 s	85.71%	1.54 s
8	88.10%	1.50 s	89.29%	1.50 s	89.29%	1.50 s

Pada pengujian diatas yaitu pengujian berdasarkan lebar *frame*, nilai koefisien MFCC dan nilai iterasi (*epoch*), pengujian dilakukan untuk mengetahui nilai akurasi maksimal dan waktu komputasi tercepat. Hal ini dilakukan untuk membuktikan bahwa akurasi dan waktu komputasi berbanding lurus, namun dalam pengujian ini hal tersebut berbanding terbalik. Ini menunjukkan bahwa sistem yang dibuat pada tugas akhir ini belum cukup baik. Sistem yang baik dan bagus adalah sistem yang memperhatikan nilai keduanya yaitu nilai akurasi maksimal dan waktu komputasi yang cepat.

- **Skenario 4** : Pengujian ketahanan *noise*



Gambar 7 Grafik Akurasi Ketahanan *Noise*



Gambar 8 Grafik Waktu Komputasi Ketahanan *Noise*

5. KESIMPULAN

Metode *Mel Frequency Cepstral Coefficient* (MFCC) sebagai ekstraksi ciri dan *Fuzzy Logic* sebagai klasifikasi dapat digunakan untuk mendeteksi tingkat usia muda, dewasa dan tua berbasiskan *speech recognition*. Parameter lebar *frame*, koefisien MFCC dan nilai iterasi (*epoch*) berpengaruh terhadap akurasi sistem. Nilai akurasi maksimal sebesar 94.05% dengan waktu komputasi rata-rata selama 1.59 *second* dihasilkan oleh lebar *frame* sebanyak 128 sampel, koefisien MFCC sebanyak 6 buah koefisien dan nilai iterasi (*epoch*) sebanyak 10 kali dengan 84 data uji. Pada sistem ini, nilai akurasi sistem berbanding terbalik dengan waktu komputasi dimana waktu komputasi tercepat yang dihasilkan adalah 0.49 *second* dengan lebar *frame* 512, koefisien MFCC sebanyak 4 buah dan nilai iterasi (*epoch*) sebanyak 10 kali dengan akurasi sebesar 76.19%. Sistem masih belum teruji dengan baik apabila diberi *noise* yang cukup besar yaitu kurang dari 40dB.

Adapun saran untuk penelitian selanjutnya adalah :

1. Menambah spesifikasi usia yang dapat diklasifikasikan
2. Implementasi pada perangkat *mobile* atau perangkat identifikasi khusus dan dapat digunakan secara *real time*.
3. Penggunaan metode ekstraksi ciri atau metode klasifikasi lain sebagai pembanding dari metode yang telah dipakai.

DAFTAR PUSTAKA

- [1] Bhaskoro, Susetyo Bagas. 2012. *"Aplikasi Pengenalan Gender Menggunakan Suara,"*. Bandung : Universitas Widyatama
- [2] Dr.Hardiwinoto. "Kategori Umur". 9 April 2015. <http://ilmu-kesehatan-masyarakat.blogspot.com/2012/05/kategori-umur.html>
- [3] Dubois, D. and H.Prade. 1980. *"Fuzzy Sets and System : Theory and Applications"*. New York : Academic Press
- [4] Dwifebrianti, Reni. *"Analisis Deteksi Tipe Suara PAda Pria Dan Wanita Menggunakan Metode Mel-Frequency Cepstral Coefficient Dan Klasifikasi Support Vector Machine Multi-Kelas One-Against-All"*. Bandung : Fakultas Elektro dan Komunikasi Institut Teknologi Telkom
- [5] *"Fuzzy Logic Toolbox™ User's Guide"* . 1995-2015 : The MathWorks, Inc.
- [6] Mulyani, Sri. 2008. *"Dasar-Dasar Sinyal Audio,"*. Direktorat Pembinaan SMK
- [7] Rahayu, N.K.I. 2014. *"Analisis dan Simulasi Sistem Penerjemah Kata Berbahasa Bali Ke Bahasa Inggris Berbasis Speech To Text Secara Real Time Menggunakan Metode Klasifikasi Hidden Markov Model"*. Bandung : Fakultas Teknik Elektro Universitas Telkom
- [8] Zadeh, L. A. 1965. *"Fuzzy Sets"*. Information and Control Vol. 8 PP. 338-353

