

IMPLEMENTASI METODE MFCC DAN DTW UNTUK PENGENALAN JENIS SUARA PRIA DAN WANITA

**Irham Sidik Permana¹, Youllia Indrawaty Nurhasanah¹, Andriana
Zulkarnain²**

¹Teknik Informatika Institut Teknologi Nasional Bandung

² Teknik Elektro Universitas Langlangbuana Bandung

irhamsidikp@gmail.com, youllia@itenas.ac.id, Andriana6970@gmail.com

ABSTRAK

MFCC (*Mel-Frequency Cepstral Coefficient*) dan DTW (*Dynamic Time Warping*) adalah suatu metode dalam mengolah suara, pada penelitian ini pengolahan suara yang dilakukan bertujuan untuk pengenalan jenis suara pria dan wanita. Penentuan jenis suara pria atau wanita biasanya dilakukan dalam penentuan kelompok paduan suara. MFCC merupakan metode untuk ekstraksi ciri, selain MFCC digunakan metode lain yaitu DTW yang merupakan metode pencocokan suara latih dan suara uji, teknik ini berguna untuk menghitung jarak antara dua data dengan pola-pola yang berbeda dan menghitung nilai dan jarak dari data tersebut. Pada penelitian ini dibangun suatu sistem yang dapat mengenali jenis suara pria dan wanita, Jenis suara pada pria dibagi menjadi *Tenor*, *Bariton* dan *Bass*, sedangkan pada wanita dibagi menjadi *Sopran*, *Mezzosopran* dan *Alto*. Hasil Penelitian yang didapat yaitu untuk tingkat akurasi pada wanita dengan jenis suara *alto* didapatkan presentase 80%, untuk tingkat akurasi jenis suara *mezzosopran* didapatkan 90%, untuk tingkat akurasi jenis suara *sopran* didapatkan 80%. Kemudian pada jenis suara pria, untuk jenis suara *bass* didapatkan tingkat akurasi sebesar 80%, untuk jenis suara *bariton* didapatkan tingkat akurasi 70%, dan untuk jenis suara *tenor* didapatkan tingkat akurasi sebesar 60%.

Kata kunci : *Pengenalan suara, Jenis suara pria dan wanita, MFCC, DTW.*

ABSTRACT

MFCC (Mel-Frequency cepstral coefficient) and DTW (Dynamic Time Warping) is a method of processing voice, in this case the processing voice is detect men and women voice type. Determining the type of man or woman voice is usually use in deciding a choir voice. MFCC is a method for feature extraction, a technique used in MFCC taken based on the human auditory approach, DTW method is for matching practiced voice and test voice, this technique is useful for calculating the

distance between two data with different patterns and calculate the value and range of the data. In this research want built a system that can identify male and female voice type, voice type in men divided into Tenor, Bariton and Bass, while the women divided into soprano, Mezzosopran and Alto. Results obtained to the level of accuracy in women with type alto obtained percentage of 80%, to the level of accuracy the type of mezzosopran voice obtained 90%, to the level of accuracy the type of soprano obtained 80%. For a man voice with type bass voice obtained accuracy rate of 80%, for bariton voice obtained accuracy rate of 70%, and for type of tenor voice obtained accuracy rate of 60%.

Kata kunci : Voice recognition , Type of male and female voice, MFCC, DTW.

1. PENDAHULUAN

1.1 Latar Belakang

Pengenalan suara adalah suatu pengembangan teknik untuk mengenali dan menterjemahkan masukan suara yang diucapkan oleh perangkat teknologi^[1]. Pada saat ini, sistem pengenalan ucapan menjadi sesuatu hal yang dapat memudahkan interaksi antara manusia dan perangkat teknologi yang ada.

Suara merupakan bagian dari manusia yang dapat dibedakan. Banyak hal-hal yang berkaitan dengan suara, salah satunya adalah pada bidang musik. Dalam teori musik, manusia memiliki jenis suara yang berbeda-beda yaitu terdapat suara tinggi, sedang, dan rendah baik pria maupun wanita. Jenis suara pada pria dibagi menjadi *Tenor*, *Bariton* dan *Bass*, sedangkan pada wanita dibagi menjadi *Sopran*, *Mezzosopran* dan *Alto*^[3]. Tetapi sebagian besar orang yang tidak memiliki ilmu mengenai musik mengalami kesulitan dalam menentukan jenis suara mereka. Dalam menentukan jenis suara biasanya dilakukan dengan cara manual menggunakan bantuan dari seorang ahli atau pelatih vokal dengan bantuan alat musik seperti piano. Penentuan jenis suara pria ataupun wanita biasanya dilakukan dalam penentuan kelompok paduan suara. Namun penentuan dengan cara ini memiliki ketergantungan pada pelatih vokal dan alat musik. Untuk itu dibuatlah sistem yang dapat mengenal jenis-jenis suara untuk mempermudah penentuan jenis suara pada pria dan wanita.

Pada proses pengenalan jenis suara pada pria dan wanita dibutuhkan suatu metode, salah satunya adalah MFCC (Mel Frequency Cepstrum Coefficients), metode ini digunakan untuk melakukan ekstraksi ciri (feature extraction), yaitu untuk mendapatkan suatu parameter dan informasi mengenai ciri dari suara seseorang^[6]. Sedangkan untuk pencocokan pola suara latih terhadap suara yang akan diuji menggunakan metode DTW (*Dinamic Time Warping*), Teknik DTW ditunjukkan untuk mengatasi perbedaan waktu. Proses perbedaan waktu yang dimaksud adalah waktu pada saat proses perekaman pengujian dengan waktu pada *template* sinyal referensi yang tersedia, prinsip dasarnya adalah dengan memberikan sebuah rentang '*steps*' dan digunakan untuk mempertemukan lintasan yang menunjukkan *local match* (kemiripan) terbesar yang selanjutnya akan dipilih *best-matching template* ^[2].

Pada penelitian ini diterapkan metode MFCC untuk mendapatkan ciri dari suara dan DTW digunakan untuk mencocokkan suara uji dan data suara latih yang telah tersimpan sebelumnya. Melalui penelitian ini dibangun sebuah aplikasi yang berfungsi untuk mengetahui jenis suara yang terdapat pada pria dan wanita serta mengetahui tingkat akurasi yang didapat dari metode MFCC dan DTW tersebut.

1.2 Rumusan Masalah

Permasalahan yang dibahas pada penelitian ini, yaitu:

1. Bagaimana membuat sistem yang dapat mendeteksi jenis suara pada pria dan wanita.
2. Bagaimana menerapkan metoda MFCC sebagai metode untuk ekstraksi ciri dan metode DTW sebagai pencocokan suara latih dan suara uji.

1.3 Tujuan

Tujuan dilakukannya penelitian ini adalah untuk membuat aplikasi yang dapat mendeteksi jenis suara pada pria dan wanita, dengan menerapkan MFCC sebagai metode ekstraksi ciri dan DTW sebagai pencocokan suara latih dan suara uji untuk dapat menganalisa jenis suara pada pria dan wanita.

1.4 Batasan Masalah

Batasan masalah pada penelitian ini adalah sebagai berikut :

1. Jenis suara yang dideteksi adalah *Sopran, Mezzosopran* dan *Alto* pada wanita, sedangkan pada pria adalah *Tenor, Bariton* dan *Bass*.
2. Durasi dari suara yang direkam sebagai data uji dan data latih yaitu huruf vokal /a/ berdurasi 2 detik dengan frekuensi sample 16000 Hz.
3. Data uji diambil dari suara ahli vokal yang dilakukan pengujian terhadap aplikasi yang dibangun.

2.1 METODOLOGI PENELITIAN

1.1 Studi Literatur

Pada tahap ini informasi dan referensi mengenai jenis-jenis suara pada pria dan wanita, konsep *Mel-Frequency Cepstral Coefficient (MFCC)* dan *Dynamic Time Warping (DTW)*, khususnya mengenai cara mendeteksi jenis suara dengan menggunakan MFCC serta DTW, dikumpulkan dan dijadikan landasan pada penelitian ini.

1.2 Subjek penelitian

Subjek penelitian pada penelitian ini adalah enam puluh data suara uji untuk 10 suara dari masing masing jenis suara nya seperti Bass, Bariton, Tenor, Alto, Mezzosopran, Sopran dan digunakan MFCC untuk mendapatkan nilai dari vektor suara tersebut yang berguna untuk mendapatkan suatu parameter dan informasimengenai ciri dari suara uji, lalu digunakan juga DTW untuk mencocokkan antara vektor ciri dari suara uji dengan suara latih.

1.3 Studi Pustaka

Pada tahap ini dilakukan pencarian mengenai pustaka atau penelitian-penelitian yang telah dilakukan sebelumnya yang berhubungan dengan topik atau masalah yang sedang diteliti yang berguna untuk membantu penulis didalam melakukan penelitian ini.

2. Metodologi Pembangunan Sistem

Adapun metodologi pembangunan sistem pada aplikasi pengenalan jenis suara dengan ekstraksi MFCC dan pencocokan DTW yang dibangun digunakan metodologi *prototype* seperti dapat dilihat pada Tabel 1.

Tabel 1. Metodologi Pembangunan sistem

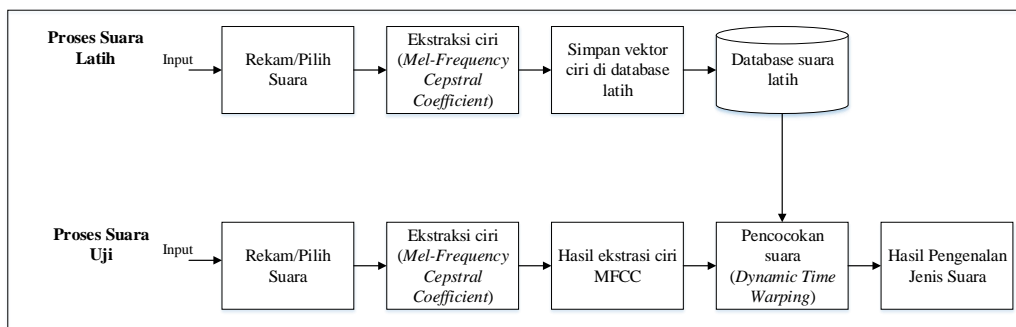
No	Fase	Tahap
1	Analisis	a. Melakukan analisis terhadap kebutuhan sistem dan fungsionalitas sistem beserta kerja sistem.
2	<i>Design</i>	a. Menjelaskan proses perancangan dengan <i>flowchart</i> . b. Menentukan design permodelan <i>layout prototype</i> .
3	Pengembangan <i>Prototype</i>	Mengembangkan bentuk <i>prototype</i> dari model hasil tahapan <i>design</i> ke dalam bentuk bahasa pemrograman.
4	Pengujian	Melakukan bentuk pengujian dengan menggunakan tabel uji.
5	Evaluasi	Melakukan evaluasi dan perbaikan pada sistem secara keseluruhan berdasarkan hasil pengujian yang telah dilakukan

3. ANALISIS DAN PEMBAHASAN

Aplikasi pengenalan jenis suara yang dibangun memiliki beberapa tahap, proses awal adalah dimasukan data suara latih yang dijadikan sebagai *database* dengan menggunakan *mel-frequency cepstral coefficient* (MFCC). MFCC merupakan salah satu teknik ekstraksi ciri (*feature extraction*) pada bidang *speech recognition* yang dilakukan dengan cara mengubah sinyal suara menjadi beberapa parameter yang dapat dianalisis.

Pada proses pengambilan data uji dilakukan perekaman suara dengan bentuk (.wav) dan dilakukan pencocokan antara suara latih dan suara uji dengan menggunakan DTW (*Dynamic Time Warping*), teknik ini ditujukan untuk mengakomodasi perbedaan waktu antara proses perekaman saat pengujian

dengan yang tersedia pada *template* sinyal referensi. Prinsip dasarnya adalah dengan memberikan sebuah rentang '*steps*' pada sebuah *frame-frame* waktu dalam *sample* dan *frame-frame* waktu dalam *template* kemudian digunakan untuk mempertemukan lintasan yang menunjukkan *local match* terbesar (kemiripan) antara time *frame* yang lurus, hasil dari pencocokan antara suara latih dan suara uji akan didapatkan jenis suara dari *user*. Gambaran umum dari sistem dapat dilihat pada Gambar 1.



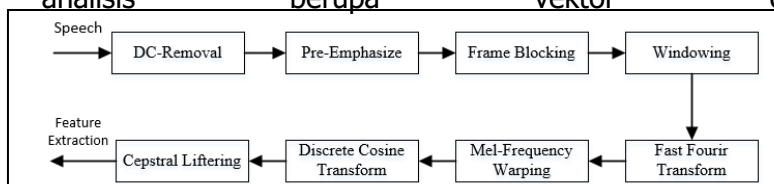
Gambar 1. Gambaran umum sistem

Berikut merupakan proses utama untuk pengenalan jenis suara pada aplikasi yang dibangun yang terdiri dari proses data suara latih dan proses suara uji.

3.1 Proses Suara Latih

Pada proses ini berguna untuk menyimpan sampel data pola jenis suara seperti *Tenor, Bariton, Bass* untuk pria, dan *Sopran, Mezzosopran, Alto* untuk wanita. Proses untuk suara latih adalah sebagai berikut:

- a. Suara dari *user* yang dijadikan sebagai suara latih didapatkan dari proses perekaman atau suara tersebut dipilih dari direktori. Suara tersebut memiliki bentuk (.wav) dengan frekuensi sample 16000 Hz, dari suara tersebut didapatkan nilai berupa *array* suara latih yang akan dilakukan tahap ekstraksi ciri dengan menggunakan MFCC. Contoh sample data *array* dari suara *input user* adalah sebagai berikut : (10, 11, 13, 14, 16, 18, 20, 9, 6, 4).
- b. Proses ekstraksi ciri dilakukan dengan menggunakan *mel-frequency cepstral coefficient* (MFCC) untuk mendapatkan beberapa parameter yang dapat di analisis berupa vektor ciri, tahap



Gambar 2. Proses Mel Frequency Cepstral Coefficient

Adapun penjelasan dari proses MFCC pada Gambar 2 adalah sebagai berikut:

1. *DC Removal* ditujukan untuk menghitung rata-rata dari data sampel suara dan mengurangi nilai setiap sampel suara dengan nilai rata-rata tersebut, tujuannya adalah membuang data-data yang tidak dibutuhkan di dalam proses suara latih. Proses DC Removal ditampilkan pada Persamaan 1

$$y[n] = x[n] - \bar{x}, 0 \leq n \leq N-1 \quad \dots\dots\dots(1)$$

Dimana:

$y[n]$ = Sample Signal hasil *DC Removal* \bar{x} = Nilai rata-rata sample signal asli
 $x[n]$ = Sample signal asli N = Panjang signal

Untuk menghitung *DC Removal* pada sampel data sinyal seperti:

(10, 11, 13, 14, 16, 18, 20, 9, 6, 4)

Hitung rata-rata sampel sinyal uji : $\bar{x} = \frac{10+11+13+14+16+18+20+9+6+4}{10} = 12.1$

Setelah nilai rata-rata diketahui, kurangkan nilai sinyal awal dengan hasil *DC Removal* sehingga nilai sinyal menjadi :

$$y_0 = 10 - 12,1 = - 2,1$$

lakukan cara yang sama untuk nilai sample lain sehingga data sinyal setelah dilakukan *DC Removal* adalah :

(- 2,1 , - 1,1 , 0,9 , 1,9 , 3,9 , 5,9 , 7,9 , - 3,1 , - 6,1 , - 8,1)

2. *Pre emphasize* dilakukan untuk mengurangi *noise* pada suara masukan, sehingga tingkat akurasi dari proses ekstraksi ciri dapat ditingkatkan. Proses *pre emphasize* ditampilkan pada Persamaan 2.

$$y[n] = s[n] - a.s[n - 1], 0.9 \leq a \leq 1.0 \quad \dots\dots\dots(2)$$

Dimana:

$y[n]$ = *signal* hasil *pre-emphasize filter*
 $s[n]$ = *signal* sebelum *pre-emphasize filter*
 a = nilai *alpha*

Dalam melakukan penghitungan *pre emphasize* dengan data hasil dari *DC Removal* seperti (- 2,1 , - 1,1 , 0,9 , 1,9 , 3,9 , 5,9 , 7,9 , - 3,1 , - 6,1 , - 8,1) dengan $\alpha = 0,97$ adalah sebagai berikut:

$$Y_0 = -2.1$$

$$Y_1 = (-1,1 - (-2,1 * 0,97)) = 0,937$$

$$Y_2 = (0,9 - (-1,1 * 0,97)) = 1,967$$

Data sinyal baru adalah data sinyal sebelum proses *pre emphasis* ditambah dengan data hasil *pre emphasis* di atas. Sehingga sinyal setelah *pre emphasis* :

$$N_n = N_n + Y_n$$

$$N_0 = -2,1 + (-2,1) = -4,2$$

$$N_1 = -1,1 + (0,937) = -0,163$$

Sehingga didapatkan nilai sinyal setelah *pre emphasis* adalah :

$$(-4,2 , -0,163 , 2,867 , 2,927 , 5,957 , 8,017 , 10,067 , -13,863 , -9,193 , -10,283)$$

3. *Frame blocking* adalah suatu proses untuk membagi sampel suara menjadi beberapa *frame* atau slot. Proses *frame blocking* ditampilkan Persamaan 3.

$$\boxed{\text{Jumlah frame} = Ts/M} \dots\dots\dots(3)$$

Dimana :

T_s = Durasi pengambilan suara (*ms*)
 M = Panjang *frame* (*ms*)

Maka untuk menghitung $T_s = 0.2$, $M = 16000$ adalah $\text{frame} = 16000/0,02 = 8000$

4. Proses *windowing* adalah proses untuk mengurangi kesenjangan (*discontinuitas*) sinyal pada awal dan akhir bingkai setelah proses *frame blocking*. Seperti pada Persamaan 4.

$$\boxed{x(n) = x_i(n)w(n)} \dots\dots\dots(4)$$

Dimana:

$x(n)$ = nilai *sample signal* hasil *windowing*
 $x_i(n)$ = nilai *sample signal* dari *frame signal* ke i
 $w(n)$ = fungsi *window*

Untuk menghitung *windowing* prosesnya adalah sebagai berikut:

$$Wn = 0.54 - 0.46 \cos \frac{2 \times 3.14 \times n}{8000 - 1} = 0.08$$

Sehingga,

$$X_0 = -4,1 * 0.08 = - 0,33$$

$$X_1 = -0,163 * 0.08 = 0,01$$

Dengan cara yang sama lakukan pada titik yang lain, dan diperoleh nilai sebagai berikut:

$$(- 0,33 , 0,01 , 0,23 , 0,23 , 0,48 , 0,64 , 0,80 , -1,11 , -0,73 , -0,82)$$

5. *Fast Fourier Transform* (FFT) berfungsi untuk mengubah domain waktu menjadi domain frekuensi.

Proses FFT ditampilkan pada persamaan 5.

$$f(n) = \sum_{k=0}^{N-1} y_k e^{-2\pi jkn/N}, n=0,1,2,\dots,N-1 \dots\dots\dots(5)$$

Dimana :

$$f(n) = \text{Frekuensi} \quad k = 0,1,2,\dots,(N-1)$$

$$N = \text{Jumlah sample pada masing-masing frame} \quad j = \text{Bilangan}$$

imajiner ($\sqrt{-1}$)

Untuk menghitung FFT pada data sinyal hasil *windowing* seperti $(- 0,33 , 0,01 , 0,23 , 0,23 , 0,48 , 0,64 , 0,80 , -1,11 , -0,73 , -0,82)$ adalah sebagai berikut :

$$F0 = \left[-0.33 \left(\cos \left(\frac{2\pi \cdot 0 \cdot 0}{10} \right) \right) \right] - j \sin \left(\cos \frac{2\pi \cdot 0 \cdot 0}{10} \right) +$$

$$\left[-0.01 \left(\cos \left(\frac{2\pi \cdot 0 \cdot 1}{10} \right) \right) \right] j \sin \left(\cos \frac{2\pi \cdot 0 \cdot 1}{10} \right) + \dots + \left[-0.82 \left(\cos \left(\frac{2\pi \cdot 0 \cdot 9}{10} \right) \right) \right] -$$

$$j \sin \left(\cos \frac{2\pi \cdot 0 \cdot 9}{10} \right) = -0.16 + 0j = -0.16$$

Dengan cara yang sama lakukan kepada data sinyal lainnya. Sehingga diperoleh data sinyal hasil FFT adalah :
 $(-0.16 , -0.32 , -0.48 , -0.64 , -0.8 , -0.96 , -1.12 , -1.28 , -1.44 , 1.6)$

6. *Mel Frequency Warping* dilakukan menggunakan *filterbank* seperti yang ditunjukkan pada Persamaan 6.

$$Mel f = \frac{2595 * \log_{10} \left(1 + \frac{f}{700} \right)}{\frac{Si}{2}} \dots\dots\dots(6)$$

Dimana :

$$Si = \text{Sinyal awal hasil FFT}$$

$$f = f_0 - f_n$$

Untuk data sinyal seperti $(-0.16 , -0.32 , -0.48 , -0.64 , -0.8 , -0.96 , -1.12 , -1.28 , -1.44 , 1.6)$ perhitungan *mel frequency warping* nya adalah sebagai berikut:

$$H_0 = \left(\frac{2595 \cdot \log(1+1000/700)}{-0.16/2} \right) = -3124.95$$

Maka penghitungan *filterbank* untuk Y_0 sebagai berikut :

$$Y_0 = -0.16 \cdot -3124.95 = 499.99$$

7. DCT adalah proses untuk menghitung *mel spectrum* sehingga menghasilkan representasi yang baik dari spektral suara. Proses DCT ditampilkan pada Persamaan 7

$$C_n = \sum_{k=1}^K (\log S_k) \cos \left[n \left(k - \frac{1}{2} \right) \frac{\pi}{K} \right]; \quad n = 1, 2, \dots, K \quad (7)$$

Dimana:

S_k = Keluaran dari proses *filterbank* pada *index* k

K = Jumlah koefisien

Untuk perhitungan DCT pada nilai hasil *filterbank* = 499.99, Koef (k) = 10, untuk n = 0 maka,

$$C_0 = \log(499.99) \cos[0(0-1/2)3.14/10] + \log(499.99) \cos[0(1-1/2)3.14/10] + \dots$$

$$\dots + \log(499.99) \cos[0(9-1/2)3.14/10] = 26.987$$

Jadi nilai DCT untuk $C_0=26.987$, lakukan cara yang sama untuk data sinyal yang lain.

8. Proses *cepstral liftering* merupakan langkah terakhir dari proses utama MFCC berfungsi untuk meningkatkan kualitas pengenalan suara. Proses *cepstral liftering* ditampilkan pada Persamaan 8.

$$W[n] = \begin{cases} 1 + \frac{L}{2} \sin\left(\frac{n\pi}{L}\right) & n = 1, 2, \dots, L \\ 0 & \end{cases} \quad (8)$$

Dimana:

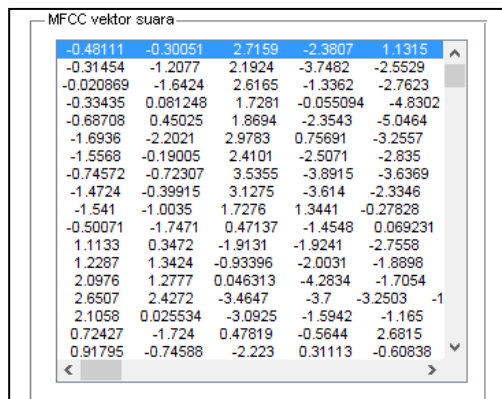
L = Jumlah *cepstral coefficients*

N = *Index* dari *cepstral coefficients*

Hasil perhitungan untuk nilai DCT= 26.987 maka, untuk *cepstral liftering* pada W_0 adalah sebagai berikut:

$$W_0 = 26.987 \cdot 10/2 \cdot \sin(3.14/10) \\ = 0.7392$$

Jadi hasil *cepstral liftering* pada $W_0=0.7392$, lakukan cara yang sama untuk data sinyal yang lain dan dari perhitungan *cepstral liftering* didapatkan suatu vektor ciri yang berguna untuk pencocokan antara suara latih dan suara uji seperti ditunjukkan pada Gambar 3.



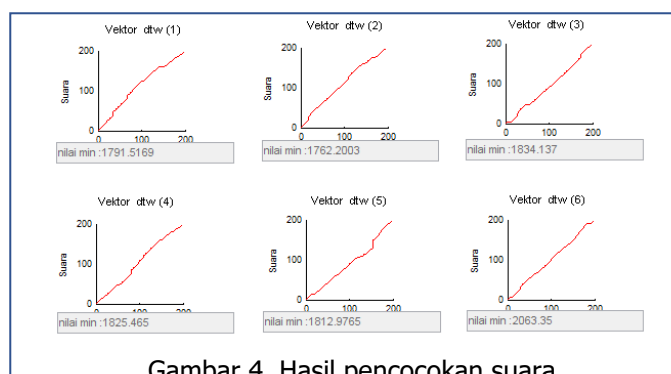
Gambar 3. Hasil ekstraksi suara latih

- c. Hasil dari proses ekstraksi didapatkan suatu vektor ciri yang disimpan pada *database* yang telah ditentukan.
- d. *Database* suara latih akan digunakan sebagai bahan referensi untuk pencocokan data uji dan data latih dalam proses pengenalan jenis suara pada pria dan wanita.

3.2 Proses suara uji

Pada proses ini dilakukan pengenalan terhadap suara *input* dari *user*. Proses untuk suara uji adalah sebagai berikut:

- a. Suara dari *user* yang dijadikan sebagai suara uji didapatkan dari proses perekaman atau suara tersebut dipilih dari direktori. Suara tersebut memiliki bentuk (.wav) dengan frekuensi sample 16000 Hz.
- b. Pada tahap ekstraksi ciri data uji digunakan *mel-frequency cepstral coefficient* (MFCC) sama seperti pada proses ekstraksi suara latih pada Gambar 2.
- c. Proses MFCC pada suara uji berfungsi untuk mendapatkan beberapa parameter yang dapat di analisis berupa vektor ciri yang akan dicocokkan dengan hasil ekstraksi ciri dari suara latih.
- d. Vektor ciri dari suara input *user* dicocokkan dengan *database* yang sebelumnya telah tersimpan pada *database* suara latih. Pencocokan digunakan dengan menggunakan DTW yaitu vektor uji dicocokkan dengan vektor suara latih dan didapatkan nilai minimum yang menunjukkan jenis suara dari masukan *user* terhadap aplikasi yang dibangun seperti dapat dilihat pada Gambar 4.



Gambar 4. Hasil pencocokan suara

Pada Gambar 4 ditampilkan sampel enam grafik dari suara yang cocok dengan pengenalan jenis suara dari *user*, grafik tersebut merupakan pencocokan hasil DTW antara vektor suara latih dan vektor suara uji, serta ditampilkan juga nilai minimum yang cocok dengan suara tersebut. Proses DTW ditampilkan pada Persamaan 9.

$$D(U, V) = \gamma(m, n)$$

$$\gamma(m, n) = d_{base}(u_i, v_j) + \min \begin{cases} \gamma(i-1, j) \\ \gamma(i-1, j-1) \\ \gamma(i, j-1) \end{cases} \dots\dots\dots(9)$$

Adapun contoh perhitungan jarak DTW untuk mendapatkan nilai minimum dari data latih dan data uji dapat dilihat seperti pada Tabel 2 . Pada tabel 3 ditunjukkan jarak terjumlahkan (*cumulative distance*) antara 2 vektor yaitu $U=\{2,5,2,5,3\}$, $V=\{0,3,6,0,6,1\}$ dimana U merupakan *Feature* data dan V merupakan *Reference* data.

Tabel 2. Contoh Perhitungan nilai DTW

	=Kolom1	Kumulatif	=Kumulatif	
$(0-2)^2$	4	0+4	4	→Minimum
$(0-5)^2$	25	4+25	29	
$(0-2)^2$	4	29+4	33	
$(0-5)^2$	25	33+25	58	
$(0-3)^2$	9	58+9	67	

Tabel 3. Cummulative distance 2 vektor

		v					
		0	3	6	0	6	1
u	2	4	5	21	25	41	42
	5	29	8	6	31	26	42
	2	33	9	22	10	26	27
	5	58	13	10	35	11	27
	3	67	13	19	19	20	15

- e. Dari pencocokan antara suara latih dan suara uji didapatkan hasil jenis suara dari *user* tersebut yaitu *Tenor, Bariton, Bass, Sopran, Mezzosopran, Alto*.

4.3 Pengujian Sistem

Pengujian pada aplikasi yang dibangun dilakukan sebanyak 60 suara uji untuk 10 kali pada masing-masing dari jenis suara yaitu *alto, mezzosopran, sopran, bass, bariton, dan tenor*. Untuk menghitung tingkat validasi pada penelitian ini dihitung dalam (%) seperti pada Persamaan berikut :

$$\frac{\text{Hasil Yang Sesuai/Benar}}{\text{Total Seluruh Uji Coba}} \times 100 \%$$

Tabel 4. Hasil pengujian suara *alto*

Tabel 5. Hasil pengujian *mezzosopran*

No	file	terdeteksi	Frekuensi	Nilai dtw
1	Alto1.wav	Alto	262.950 Hz	1194.388
2	Alto2.wav	Alto	262.295 Hz	1194.388
3	Alto3.wav	Mezzo	262.295 Hz	1941.443
4	Alto4.wav	Alto	262.295 Hz	836.014
5	Alto5.wav	Alto	258.006 Hz	671.163
6	Alto6.wav	Alto	258.065 Hz	1310.85
7	Alto7.wav	Sopran	340.426 Hz	1761.625
8	Alto8.wav	Alto	258.065 Hz	554.311
9	Alto9.wav	Alto	262.295 Hz	648.959
10	Alto10.wav	Alto	258.065 Hz	323.394

No	File	Terdeteksi	Frekuensi	Nilai dtw
1	Mezzo1.wav	Mezzosopran	470.588 Hz	1933.694
2	Mezzo2.wav	Sopran	363.636 Hz	1941.320
3	Mezzo3.wav	Mezzosopran	380.952 Hz	1982.229
4	Mezzo4.wav	Mezzosopran	380.952 Hz	2024.509
5	Mezzo5.wav	Mezzosopran	380.952 Hz	1992.506
6	Mezzo6.wav	Mezzosopran	363.636 Hz	2107.368
7	Mezzo7.wav	Mezzosopran	363.636 Hz	2083.172
8	Mezzo8.wav	Mezzosopran	470.588 Hz	2104.743
9	Mezzo9.wav	Mezzosopran	363.636 Hz	2086.619
10	Mezzo10.wav	Mezzosopran	258.065 Hz	1964.273

Berdasarkan Tabel 4 maka hasil dari pengujian untuk jenis suara *alto* didapatkan tingkat akurasi sebesar:

$$\frac{8}{10} \times 100\% = 80\%.$$

Sedangkan hasil dari pengujian untuk jenis suara *mezzosopran* berdasarkan Tabel 5 didapatkan tingkat akurasi sebesar:

$$\frac{9}{10} \times 100\% = 90\%.$$

Tabel 6. Hasil pengujian suara *sopran*

No	File	Terdeteksi	Frekuensi	Nilai dtw
1	Sopran1.wav	Sopran	727.273 Hz	1966.858
2	Sopran2.wav	Alto	277.273 Hz	2544.622
3	Sopran3.wav	Sopran	695.652 Hz	2229.570
4	Sopran4.wav	Sopran	695.652 Hz	2229.570
5	Sopran5.wav	Sopran	761.905 Hz	1150.348
6	Sopran6.wav	Sopran	421.053 Hz	2476.880
7	Sopran7.wav	Sopran	941.176 Hz	2509.588
8	Sopran8.wav	Mezzosopran	457.143 Hz	2093.873
9	Sopran9.wav	Sopran	941.176 Hz	1839.055
10	Sopran10.wav	Sopran	941.176 Hz	1900.455

Untuk hasil pengujian jenis suara *sopran* dapat dilihat pada Tabel 6 dan mendapatkan tingkat akurasi sebesar:

$$\frac{8}{10} \times 100\% = 80\%.$$

Tabel 7. Hasil pengujian Bass

No	file	terdeteksi	Frekuensi	Nilai dtw
1	Bass1.wav	Bass	118.519 Hz	1162.44
2	Bass2.wav	Bass	118.519 Hz	1568.501
3	Bass3.wav	Bariton	363.636 Hz	2217.977
4	Bass4.wav	Bass	106.667 Hz	1672.687
5	Bass5.wav	Bass	118.519 Hz	1091.104
6	Bass6.wav	Bass	106.667 Hz	1602.687
7	Bass7.wav	Bass	118.519 Hz	945.219
8	Bass8.wav	Bass	118.519 Hz	701.503
9	Bass9.wav	Bariton	363.636 Hz	2391.065
10	Bass10.wav	Bass	106.667 Hz	1619.498

Berdasarkan tabel 7 maka hasil dari pengujian untuk jenis suara *bass* didapatkan tingkat akurasi sebesar:

$$\frac{8}{10} \times 100\% = 80$$

Tabel 8. Hasil pengujian Bariton

No	File	Terdeteksi	Frekuensi	Nilai dtw
1	Bariton1.wav	Bariton	258.065 Hz	1762.203
2	Bariton2.wav	Bariton	262.295 Hz	1726.291
3	Bariton3.wav	Bariton	262.295 Hz	1500.941
4	Bariton4.wav	Tenor	258.065 Hz	1537.663
5	Bariton5.wav	Bariton	285.714 Hz	1640.683
6	Bariton6.wav	Bass	313.725 Hz	1544.221
7	Bariton7.wav	Bariton	262.295 Hz	1628.74
8	Bariton8.wav	Bariton	262.295 Hz	1639.076
9	Bariton9.wav	Bass	258.065 Hz	1538.501
10	Bariton10.wav	Bariton	262.295 Hz	1556.305

Tabel 9. Hasil pengujian Tenor

No	file	Terdeteksi	Frekuensi	Nilai dtw
1	Tenor1.wav	Tenor	380.952 Hz	1572.002
2	Tenor2.wav	Bass	380.952 Hz	1997.839
3	Tenor3.wav	Tenor	380.952 Hz	960.275
4	Tenor4.wav	Tenor	380.952 Hz	1532.566
5	Tenor5.wav	Tenor	390.244 Hz	1590.883
6	Tenor6.wav	Bass	347.826 Hz	1784.478
7	Tenor7.wav	Tenor	390.224 Hz	1590.883
8	Tenor8.wav	Bass	367.692 Hz	1828.7529
9	Tenor9.wav	Bass	231.884 Hz	1495.434
10	Tenor10.wav	Tenor	380.952 Hz	1657.625

Untuk hasil pengujian jenis suara *bariton* dapat dilihat pada tabel 8 dan mendapatkan tingkat akurasi sebesar:

$$\frac{7}{10} \times 100\% = 70\%.$$

Untuk hasil pengujian jenis suara *tenor* dapat dilihat pada tabel 9 dan mendapatkan tingkat validasi sebesar:

$$\frac{6}{10} \times 100\% = 60\%$$

4. KESIMPULAN

Berdasarkan penelitian mengenai aplikasi untuk pengenalan jenis suara pada pria dan wanita dengan menerapkan metode *mel-frequency cepstral coefficient* (MFCC) dan *Dynamic time warping* (DTW) yang dilakukan, diperoleh kesimpulan sebagai berikut :

Aplikasi yang dibangun dapat mengenali jenis suara dengan tingkat akurasi yang berbeda-beda, yaitu untuk tingkat akurasi pada wanita dengan jenis suara *alto* berdasarkan Tabel 2 didapatkan presentase 80%, untuk tingkat akurasi jenis suara *mezzosopran* berdasarkan Tabel 3 didapatkan tingkat akurasi sebesar 90%, untuk tingkat akurasi jenis suara *sopran* berdasarkan Tabel 4 didapatkan tingkat akurasi sebesar 80%. Kemudian pada jenis suara pria, untuk jenis suara *bass* berdasarkan Tabel 5 didapatkan tingkat akurasi sebesar 80%, untuk jenis suara *bariton* berdasarkan Tabel 6 didapatkan tingkat akurasi 70%, dan untuk jenis suara *tenor* berdasarkan Tabel 7 didapatkan tingkat akurasi sebesar 60%.

DAFTAR RUJUKAN

- [1] Angga, S. (2011), *Pengenalan Ucapan dengan Ekstraksi Mel-Frequency Cepstrum Coefficients (MFCC)*. Semarang : Universitas Diponegoro.
- [2] Darma Putra. (2011), *Verifikasi Biometrika Suara Menggunakan Metode MFCC dan DTW*. Bali : Universitas Udayana.
- [3] M. N. Al-Azhar, *Audio Forensic: Theory And Analysis*. Pusat Laboratorium Forensik Polri Bidang Fisika Dan Komputer Forensik, 2011.

- [4] Manunggal, HS. 2005. *Perancangan dan Pembuatan Perangkat Lunak Pengenalan Suara Pembicara dengan Menggunakan Analisa MFCC Feature Extraction*. Surabaya.
- [5] Pressman, R.S. 2010, *Software Engineering : a practitioner's approach*. McGraw-Hill, New York.
- [6] Rabiner, Lawrence, and Biing-Hwang Juang, 1993, "*Fundamentals of Speech Recognition*". Prentice Hall, New Jersey.

- [7] Waluyanti, Sri, dkk. 2008. *Buku Direktorat PSMK Untuk Teknik Audio Video*.