

Analisis *Speaker Recognition* menggunakan Metode *Dynamic Time Warping* (DTW) Berbasis Matlab

Noor Fita Indri Prayoga, Yenni Astuti*, Catur Budi Waluyo

Departemen Teknik Elektro, Sekolah Tinggi Teknologi Adisutjipto, Yogyakarta

*email: yenni.stta@gmail.com

Abstract

Voice is one of way to communicate and express yourself. Speaker recognition is a process carried out by a device to recognize the speaker through the voice. This study designed a speaker recognition system that was able to identify speakers based on what was said by using dynamic time warping (DTW) method based in Matlab. To design a speaker recognition system begins with the process of reference data and test data. Both processes have the same process, which starts with sound recording, preprocessing, and feature extraction. In this system, the Fast Fourier Transform (FFT) method is used to extract the features. The results of the feature extraction process from the two data will be compared using the DTW method. Calculations using DTW that produce the smallest value will be determined as the output. The test results show that the system can identify the voice with the best level of recognition accuracy of 90%, and the average recognition accuracy of 80%. The results were obtained from 50 tests, carried out by 5 people consisting of 3 men and 2 women, each speaker said a predetermined word

Keyword — Speaker Recognition, Dynamic Time Warping, Fast Fourier Transform.

1. Pendahuluan

Suara adalah salah satu cara untuk berkomunikasi dan mengekspresikan diri. Secara perlahan dengan didukung oleh perkembangan teknologi, kebutuhan akan adanya sistem dan aplikasi yang mampu menganalisis sinyal suara pun semakin tinggi. Pemanfaatan aplikasi ini semakin berkembang, mulai dari sarana pembelajaran hingga bidang keamanan. Suatu upaya untuk dapat mengenali seorang individu melalui media suara disebut dengan *voice recognition*. *Voice recognition* terbagi menjadi dua jenis, yaitu *speech recognition* dan *speaker recognition*. *Speech recognition* adalah proses yang dilakukan suatu perangkat untuk mengenali kata yang diucapkan dan mencocokkannya dengan suatu pola yang tersimpan dalam suatu perangkat tertentu. *Speaker recognition* adalah proses yang dilakukan suatu perangkat untuk mengenali pembicara melalui suara.

Permasalahan yang sering ditemui dalam *speaker recognition* adalah pada saat perekaman, karena pada saat perekaman sering kali terjadi perbedaan durasi walaupun kata atau kalimat yang diucapkan sama. Sebagai akibatnya proses pencocokan antara data uji dengan data referensi menghasilkan nilai yang tidak optimal. Permasalahan tersebut dapat diselesaikan dengan metode *dynamic time warping* yang ditunjukkan untuk mengakomodasikan perbedaan waktu antara proses pengujian dan data referensi.

Berdasarkan latar belakang tersebut maka dibuatlah penelitian tugas akhir berupa analisis *speaker recognition* menggunakan metode *dynamic time warping* berbasis Matlab 2016a.

1.1 Kajian Pustaka

Penelitian oleh Rianto, Jeri (2011), dengan judul “Perangkat Lunak Pengenalan Suara (*Voice Recognition*) Untuk Absensi Karyawan dengan Menggunakan Metode *Dynamic Time Warping* (DTW)”. Pada penelitian tersebut membahas mengenai pencocokan suara untuk mengenali suara yang diucapkan. Pengujian yang dilakukan dalam penelitian tersebut dilakukan berdasarkan 5 pengujian dengan kondisi dan pengucapan intonasi yang berbeda-beda. Untuk pengujian ke-1, 2 dan 3 masing-masing dilakukan dengan cara pengenalan suara pada kondisi suasana sunyi, pengenalan suara pada kondisi suara suasana tidak sunyi (*ber-noise*) dan pengenalan suara dengan intonasi panjang. Pengujian tersebut dilakukan oleh 15 orang yang terdiri dari 10 pria dan 5 wanita. Sedangkan untuk pengujian yang ke-4 dan 5 masing-masing dilakukan dengan cara pengenalan suara oleh 15 wanita dan pengenalan suara oleh 15 pria. Pengujian tersebut menunjukkan bahwa pengujian 1, 2, 3, 4 dan 5 dengan jumlah suara yang dikenali dan presentase keakuratan masing-masing adalah sebagai berikut: 14 suara dikenali (93,3%), 0 suara dikenali (0%), 5 suara dikenali (33,3%), 9 suara dikenali (60%) dan 11 suara dikenali (73,3%).

Penelitian oleh Sabarani, Ruth Airin Lamtiur (2018), dengan judul “Identifikasi Sinyal Suara Menggunakan Metode *Fast Fourier Transform* (FFT) Berbasis Matlab”. Pada Penelitian tersebut telah berhasil merancang dan membangun suatu sistem yang mampu mengidentifikasi suatu sinyal suara menggunakan metode *fast fourier transform* dengan perangkat lunak Matlab. Sistem tersebut dirancang untuk mencari nilai maksimum untuk setiap sampel suara pada hasil sinyal FFT, kemudian membuat suatu klasifikasi suara agar nantinya dapat teridentifikasi. Dari penelitian tersebut didapatkan bahwa tingkat pengenalan suara menjadi lebih baik jika nilai *frame* lebih tinggi, pada penelitian tersebut menghasilkan tingkat pengenalan terbaik yaitu 96% pada *frame blocking* 256. Namun tingkat pengenalan suara menjadi turun ketika nilai *frame blocking* berada dibawah 128, yaitu 16, 32 dan 64. Penurunan tingkat pengenalan ini dikarenakan nilai indeks maksimum setiap suara berada di *range* yang sama.

Jurnal oleh Handoko, Destian Tri (2018), dengan judul “Aplikasi Keamanan PC dengan *Voice Recognition* Menggunakan Metode MFCC dan *Dynamic Time Warping* (DTW)”. Jurnal tersebut membahas mengenai *voice recognition* dalam sistem keamanan komputer yang digunakan untuk mengamankan kinerja, data, fungsi atau proses komputer. Tingkat keberhasilan sistem keamanan tersebut dipengaruhi oleh tingkat kebisingan atau *noise*. Hasil dari penelitian tersebut menampilkan bahwa sistem masih dapat diakses oleh 29% orang. Semakin sedikit presentase keberhasilan seseorang membuka sistem keamanan, maka sistem keamanan tersebut semakin baik.

1.2 Landasan Teori

Adapun beberapa dasar teori untuk mendukung penelitian ini adalah sebagai berikut.

A. Sinyal Suara

Sinyal adalah suatu isyarat untuk melanjutkan atau meneruskan suatu kegiatan. Biasanya isyarat ini berbentuk tanda-tanda, lampu-lampu, suara-suara, dan lain-lain. Menurut *International Telecommunication Union*, sinyal adalah suatu gejala fisika dimana satu atau lebih dari karakteristiknya melambangkan informasi. Suara merupakan fenomena fisik yang dihasilkan oleh getaran suatu benda yang berupa sinyal analog dengan amplitudo yang berubah secara kontinu terhadap waktu. Suara atau bunyi biasanya merambat melalui udara. Suara atau bunyi tidak bisa merambat melalui ruang hampa. Gelombang suara bervariasi dalam tingkatan tekanan suara (amplitudo) dan dalam frekuensi yang menyebabkan suara yang satu dengan berbeda dengan suara lainnya. Jumlah waktu yang diperlukan untuk terjadinya suatu getaran atau gelombang dinamakan periode. Sedangkan jumlah gelombang yang terjadi setiap detik dinamakan frekuensi dengan satuan Hz. Frekuensi yang dapat ditangkap oleh telinga manusia adalah antara 20 Hz sampai 20 KHz. Amplitudo adalah keras lemahnya bunyi atau tinggi

rendahnya gelombang. Amplitudo sebuah gelombang mengacu pada besarnya perubahan tekanan dan tingkat kerasnya gelombang suara. Satuan amplitudo adalah desibel (dB), bunyi dapat merusak telinga jika tingkat volumenya lebih besar dari 85 dB dan pada ukuran 130 dB akan mampu membuat hancur gendang telinga.

B. Definisi Speaker Recognition

Speaker recognition adalah proses yang dilakukan suatu perangkat untuk mengenali pembicara melalui suara. Dalam *speaker recognition* diperlukan proses referensi dan proses uji. Proses referensi diperlukan untuk semua kasus *pattern recognition*, hasil dari proses referensi akan digunakan sebagai *database* yang nantinya akan dibandingkan dengan suara uji. *Speaker recognition* dapat dilakukan dengan bergantung teks (*text dependent*) ataupun tidak bergantung teks (*text independent*). *Speaker recognition* memenuhi dua fungsi yaitu identifikasi dan verifikasi. Sistem identifikasi bertujuan untuk memecahkan identitas seseorang. Sedangkan sistem verifikasi bertujuan untuk menolak atau menerima identitas yang diklaim oleh seseorang. Dalam penelitian tugas akhir ini sistem *speaker recognition* yang dirancang hanya akan difokuskan pada sistem identifikasinya saja.

C. Preprocessing

Preprocessing adalah proses-proses awal yang dilakukan untuk memperbaiki kualitas objek. Tujuan dari *preprocessing* ini adalah menyertakan sinyal suara masukan agar lebih mudah diproses untuk pengenalan suara. Proses *preprocessing* memiliki beberapa tahap yaitu normalisasi, *frame blocking*, dan *windowing*.

1) Normalisasi

Normalisasi bertujuan agar untuk menyetarakan amplitudo suara yang terekam menjadi maksimum, agar efek kuat atau lemahnya suara yang terekam tidak mempengaruhi proses pengenalan suara. Normalisasi dilakukan dengan membagi data masukan (data sinyal suara terekam) dengan nilai *absolute* maksimum data tersebut. Perhitungan matematis untuk mencari nilai normalisasi dapat dilihat pada persamaan (1)

$$x_{norm}(n) = \frac{x_{masukan}}{\max |(x_{masukan})|} \quad (1)$$

Dimana,

$x_{norm}(n)$ = hasil data sinyal normalisasi (1,2,3,...,n)

$x_{masukan}(n)$ = data masukan dari suara yang telah direkam (1,2,3,...,n)

n = panjang sinyal.

$\max |(x_{masukan})|$ = nilai absolute maksimum sinyal masukan

2) Frame Blocking

Hasil Perekaman suara merupakan sinyal analog yang berada dalam domain waktu yang bersifat varian time. Karena sinyal suara terus mengalami perubahan akibat adanya pergeseran artikulasi dari organ produksi *vocal*, sinyal harus diproses secara *short segments*. *Frame Blocking* adalah proses pembagian suara menjadi beberapa *frame* yang nantinya dapat memudahkan dalam perhitungan dan analisis suara, satu *frame* terdiri dari beberapa sampel tergantung berapa besar frekuensi samplingnya. Panjang *frame* yang biasanya digunakan untuk pemrosesan sinyal adalah antara 20 sampai 50 milidetik. Panjang *frame* yang digunakan sangat mempengaruhi keberhasilan dalam analisa spektral. Di satu sisi, ukuran dari *frame* harus sepanjang mungkin untuk dapat menunjukkan resolusi frekuensi yang baik. Tetapi di lain sisi, ukuran *frame* juga harus cukup pendek untuk dapat menunjukkan resolusi waktu yang baik.

Selain itu, proses *frame blocking* dilakukan secara *overlapping* untuk setiap *frame*-nya. Panjang daerah *overlap* yang umum digunakan adalah kurang lebih 30% sampai 50% dari panjang *frame*. *Overlapping* dilakukan untuk menghindari hilangnya ciri karakteristik suara pada perbatasan perpotongan setiap *frame*. Untuk mengetahui jumlah *frame* yang akan digunakan dapat menggunakan persamaan matematis (2).

$$\text{Jumlah Frame} = \left(\frac{I-N}{M} \right) + 1 \quad (2)$$

Dimana:

I= total *sample*

N= *frame size*

M= *overlapping*

3) *Windowing*

Suara yang di potong-potong menjadi beberapa *frame* dapat membuat data suara menjadi *discontinue* pada awal dan akhir tiap *frame*, untuk meminimalisir hal tersebut maka perlu dilakukan *windowing*. Sampel suara yang telah dibagi menjadi beberapa *frame* perlu dijadikan suara kontinu kembali dengan cara mengalikan tiap *frame* dengan jenis *window* tertentu. Jenis *window* yang dipakai pada penelitian tugas akhir ini adalah *window hamming*. *Window Hamming* diperoleh dengan memodifikasi koefisien *window Hanning*. *Window hamming* menghasilkan *sidelobe level* yang tidak terlalu tinggi (kurang lebih -43 dB) dan *noise* yang dihasilkanpun tidak terlalu besar yaitu sebesar 1,36 dB. Persamaan matematis fungsi *window hamming* dapat dilihat pada persamaan (3) Kemudian representasi dari fungsi *window* terhadap sinyal suara input dapat dilihat pada persamaan (4).

$$w(n) = 0,52 - 0,46 \cos \left[\frac{2\pi n}{N-1} \right] \quad 0 \leq n \leq N - 1 \quad (3)$$

$$x(n) = x_i(n) \times w(n) \quad (4)$$

Dimana:

$w(n)$ = fungsi *window*

N= *frame size* atau jumlah isyarat suara dalam sebuah *frame*

$x(n)$ = nilai sampel sinyal hasil *windowing*

$x_i(n)$ = nilai sampel dari *frame* sinyal ke-i

D. Ekstraksi Ciri

Ekstraksi ciri merupakan suatu proses yang bertujuan untuk mendapatkan sederetan besaran pada bagian sinyal masukkan untuk menetapkan pola uji. Dalam proses pengenalan diperlukan pencuplikan lebih dari satu, oleh sebab itu pencuplikan yang telah dilakukan dapat dibentuk pola ujinya dengan tahapan ekstraksi ciri ini. Ekstraksi ciri mempunyai berbagai macam tipe seperti : ekstraksi ciri secara *Discrete Cosine Transform* dan ekstraksi ciri *Fast Fourier Transform* (FFT). Pada penelitian tugas akhir ini penulis menggunakan ekstraksi ciri FFT.

E. *Fast Fourier Transform*

Pada tahun 1965 sebuah makalah diterbitkan oleh *Cooley* dan *Tukey* menjelaskan algoritma yang sangat efisien untuk menerapkan DFT. Algoritma yang sekarang dikenal sebagai *Fast Fourier Transform* (FFT). Ada dua jenis algoritma FFT yaitu algoritma *Fast Fourier Transform Decimation In Time* (FFT DIT) dan algoritma *Fast Fourier Transform*

Decimation In Frequency (FFT DIF). Pada DIT, masukan disusun atau dikelompokkan menjadi kelompok ganjil dan kelompok genap. Runtun bernomor genap adalah $x(0), x(2), x(4), \dots, x(N-2)$ dan runtun bernomor ganjil adalah $x(1), x(3), x(5), \dots, x(N-1)$. Kedua runtun berisi $N/2$ -titik. Runtun genap ditandai dengan $x(2k)$ dengan $x=0$ sampai $k=N/2-1$, sedangkan runtun ganjil menjadi $x(2k-1)$. Dalam FFT menggunakan persamaan DFT yang dibagi menjadi bagian ganjil dan bagian genap. Persamaan DFT dapat dilihat pada persamaan (5) dan persamaan DFT yang dibagi menjadi bagian ganjil dan bagian genap yang dapat dilihat pada persamaan (6).

$$X(k) = \sum_{n=0}^{N-1} x(n) \cdot e^{-\frac{j2\pi nk}{N}} \quad (5)$$

Karena $W_N^2 = e^{-\frac{j2\pi 2}{N}} = e^{-\frac{j2\pi}{N/2}}$, maka $W_N^2 = W_{N/2}$. Jadi persamaan (5) menjadi:

$$X(k) = \sum_{n=0}^{\left(\frac{N}{2}\right)-1} x(2n) \cdot W_N^{nk} + W_N^{2nk} \sum_{n=0}^{\left(\frac{N}{2}\right)-1} x(2n+1) W_N^{nk} \quad (6)$$

$$X\left(k + \frac{N}{2}\right) = \sum_{n=0}^{\left(\frac{N}{2}\right)-1} x(2n) \cdot W_N^{nk} - W_N^{2nk} \sum_{n=0}^{\left(\frac{N}{2}\right)-1} x(2n+1) W_N^{nk} \quad (7)$$

Persamaan (6) dan (7) merupakan *FFT radix-2 Decimation in time* (DIT) yang mana *sequence* data dipilah menjadi dua bagian menjadi genap dan ganjil dan menggambarkan gabungan dua DFT $N/2$ data. Penggunaan sifat periodik dari fungsi kernel membuat perhitungan menjadi lebih efisien. Karena cukup mengganti tanda operasi menjadi *minus*.

Hasil perhitungan dari FFT adalah bilangan kompleks, yang terdiri dari unsur *real* dan *imaginer*. Sehingga dinyatakan dalam *magnitude* atau *phase* dengan persamaan (8) dan (9).

$$\text{Magnitude}|X| = \sqrt{(R)^2 + (I)^2} \quad (8)$$

$$\text{Phase} = \tan^{-1}\left(\frac{I}{R}\right) \quad (9)$$

F. *Dynamic Time Warping*

Tahap awal dari metode DTW adalah menghitung nilai *local distance* antar elemen dari kedua runtun. Persamaan untuk mencari *local distance*, *Local distance* dapat dihitung dengan mencari nilai *absolute* dari selisih kedua runtun data dengan menggunakan persamaan (10).

$$c_{i,j} = |x_i - y_j|, \quad i=1:m \text{ dan } j=1:n \quad (10)$$

Kemudian langkah selanjutnya merupakan mencari point-point yang harus dilewati untuk menemukan *optimal warping path*, yang dimulai dengan baris pertama atau *vertical move* ($1, j$), kolom pertama atau *horizontal move* ($i, 1$), lalu elemen lainnya atau *diagonal move* (i, j) dengan menggunakan persamaan berikut:

a. Baris Pertama

$$D(1, j) = |c(x_1, y_j)| + D(i, j - 1) \quad (11)$$

b. Kolom Pertama

$$D(i, 1) = |c(x_k, y_1)| + D(i - 1, j) \quad (12)$$

c. Elemen Lainnya

$$D(i, j) = |c(x_i, y_j)| + \min \left\{ \begin{array}{l} D[i-1, j-1] \\ D[i-1, j] \\ D[i, j-1] \end{array} \right\} \quad (13)$$

Nilai-nilai tersebut nantinya akan membentuk *optimal warping path*, dimana nilai terakhir pada pojok kanan bawah merupakan hasil perhitungan DTW berupa selisih nilai antara suara referensi dan suara yang diujikan. Setelah mendapatkan nilai DTW, *optimal warping path* dapat ditentukan yang dimulai dari *point* paling akhir melangkah hingga ke *point* awal. Contoh perhitungan menggunakan DTW dapat dilihat pada Gambar (1).

	-2	10	-10	15	-13	20	-5	14	2
3	5	12	25	37	53	70	78	89	90
-13	16	28	15	43	37	70	78	105	104
14	32	20	39	16	43	43	62	62	74
-7	37	37	23	38	22	49	45	66	71
9	48	38	42	29	44	33	47	50	57
-2	48	50	46	46	40	55	36	52	54

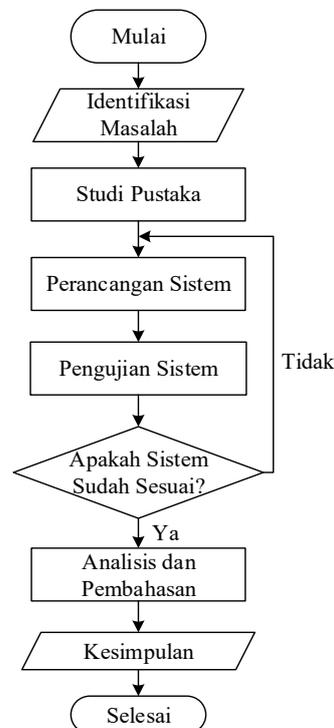
Gambar 1. Contoh Perhitungan Ekstraksi Ciri Menggunakan Metode Dynamic Time Warping

2. Metodologi Penelitian

Pada bagian ini akan dibahas mengenai alur penelitian dan perancangan sistem. Adapun metodologi dalam penelitian ini adalah sebagai berikut.

2.1 Alur Penelitian

Alur penelitian yang digunakan sebagai dasar dalam melaksanakan penelitian, ditunjukkan berdasarkan diagram alir pada Gambar 2.



Gambar 2. Diagram alir penelitian

Berdasarkan data pada Tabel 1 dapat diketahui tingkat persentase akurasi pengenalan suara menggunakan persamaan sebagai berikut:

$$\% \text{Akurasi Pengenalan} = \frac{\sum \text{Data yang dikenali}}{\sum \text{Total data yang diujikan}} \times 100\%$$

$$\% \text{Akurasi Pengenalan Suara Dicky} = \frac{8}{10} \times 100\% = 80\%$$

$$\% \text{Akurasi Pengenalan Suara Fita} = \frac{9}{10} \times 100\% = 90\%$$

$$\% \text{Akurasi Pengenalan Suara Nando} = \frac{8}{10} \times 100\% = 80\%$$

$$\% \text{Akurasi Pengenalan Suara Rio} = \frac{9}{10} \times 100\% = 90\%$$

$$\% \text{Akurasi Pengenalan Suara Titis} = \frac{8}{10} \times 100\% = 80\%$$

$$\% \text{Rata - rata pengenalan suara} = \frac{80 + 90 + 80 + 90 + 80}{5} \times 100\% = 84\%$$

Tabel 2. Hasil Akurasi Pengenalan Suara

No	Nama Pembicara	Jumlah Pengujian	Jumlah Sinyal Suara yang Dikenali	Akurasi Pengenalan (%)
1	Dicky	10	8	80%
2	Fita	10	9	90%
3	Nando	10	8	80%
4	Rio	10	9	90%
5	Titis	10	8	80%
Rata-Rata Pengenalan				84%

4. Penutup

Kesimpulan yang didapat dari hasil penelitian ini adalah:

1. Untuk merancang sistem *speaker recognition* yang dapat mengidentifikasi suara dimulai dengan proses data referensi dan data uji. Data referensi dan data uji mempunyai proses yang sama, yaitu dimulai dengan perekaman suara, *preprocessing*, dan ekstraksi ciri FFT. Hasil dari proses ekstraksi ciri FFT kedua data tersebut akan dibandingkan menggunakan metode DTW. Perhitungan menggunakan DTW yang menghasilkan nilai terkecil akan ditentukan sebagai hasil keluaran, yang akan ditampilkan berupa nama pembicara.
2. Sistem *speaker recognition* yang dibuat dengan menggunakan metode *dynamic time warping* menghasilkan tingkat akurasi pengenalan rata-rata sebesar 84% yang diperoleh dari 50 kali pengujian oleh 5 pembicara yang mengucapkan kata yang sama, dengan tingkat akurasi pengenalan terbaik sebesar 90%.

Daftar Pustaka

- [1] Atmawati, Monica Jatu Tri., 2016, *Pengenalan suara instrumen musik menggunakan analisis spektrum*, Universitas Sanata Dharma. Yogyakarta.
- [2] Dinata, Candra., Puspitaningrum, Diyah., Ernawati., 2017, *Implementasi Teknik Dynamic Time Warping (DTW) Pada Aplikasi Speech To Text*, Program Studi Teknik Informatika Universitas Bengkulu, Bengkulu.
- [3] Gunawan, Dadang., Juwono, Filbert Hilman., 2012, *Pengolahan Sinyal Digital Dengan Pemrograman Matlab*, Graha ilmu, Yogyakarta.
- [4] Handoko, Destian Tri., 2018, *Aplikasi Keamanan PC dengan Voice Recognition Menggunakan Metode MFCC dan Dynamic Time Warping (DTW)*, Universitas Nusantra PGRI Kediri, Kediri.
- [5] Hapsari, Jenny P., 2011, *Aplikasi Pengenalan Suara Dalam Pengaksesan Sistem Informasi Akademik*. Universitas Diponegoro, Semarang.
- [6] Kurniawan, Irwan., 2009, *Pengolahan Sinyal*, Politeknik Jambi, Jambi.
- [7] Kurniawan, Wawan., 2016, *Identifikasi Speech Recognition Manusia dengan Menggunakan Average Energy dan Silent Ratio Sebagai Feature Extraction Suara pada Komputer*, Universitas Jambi, Jambi.
- [8] Mathworks, 2019, *Dynamic Time Warping (DTW)*, Online.
https://www.mathworks.com/help/signal/ref/dtw.html?s_tid=srchtitle, diakses pada tanggal 15 Mei 2019
- [9] Mathworks, 2019, *Hamming Window*, Online.
https://www.mathworks.com/help/signal/ref/hamming.html?searchHighlight=Hamming&s_tid=doc_srchttitle, diakses pada tanggal 15 Mei 2019
- [10] Mulyani, Rini., 2017, *Pemetaan Dan Analisis Tipe Suara Manusia Menggunakan Fast Fourier Transform (FFT) Studi Kasus Ukm Psm Unila*, Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Lampung, Bandar Lampung.
- [11] Munir, Rinaldi., 2004, *Pengolahan Citra Digital Dengan Pendekatan Algoritmik*, Penerbit Informatika, Bandung.
- [12] Nurlaily, 2009, *Pencocokan Pola Suara Dengan Algoritma FFT Dan DC*, Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Sumatera Utara, Medan.
- [13] Permana, Irham Sidik., Nurhasanah, Youllia Indrawaty., Zulkarnain, Andriana., 2018, *Implementasi Metode MFFC Dan DTW Untuk Pengenalan Jenis Suara Pria Dan Wanita*, Teknik Informatika Institut Teknologi Nasional Bandung. Bandung
- [14] Rianto, Jeri., 2011, *Perangkat Pengenalan Suara (Voice Recognition) Untuk Absensi Karyawan Dengan Menggunakan Metode Dynamic Time Warping (DTW)*, Universitas Komputer Indonesia, Bandung.
- [15] Santoso, Tri Budi., Octavianto Hary., Huda, Miftahul., 2010, *Praktikum pengolahan sinyal digital*, Institut Teknologi Sepuluh November, Surabaya.
- [16] Setyawan, Yohanes Robby., 2014, *Pengenalan Ucapan Angka Secara Real Time Menggunakan Ekstraksi Ciri FFT dan Fungsi Similaritas Kosinus*, Universitas Sanata Dharma, Yogyakarta.
- [17] Sibarani, Ruth Airin Lamtiur. 2018. *Identifikasi Sinyal Suara Menggunakan Metode Fast Fourier Transfoem (FFT) Berbasis Matlab*. Medan. Universitas Sumatera Utara, Medan.
- [18] Sukono, Alexander., 2015, *Algoritma Dynamic Programming dalam Pengenalan Suara*, Program Studi Teknik Informatika Sekolah Teknik Elektro dan Informatika Institut Teknologi Bandung, Bandung.
- [19] Triansyah, Ersya. Indrawaty, Youllia., 2017, *Implementasi Metode Pattern Recognition Untuk Pengenalan Ucapan Huruf Hijaiyyah*, Institut Teknologi Nasional, Malang.