

# Sistem pengenalan emosi berbasis suara menggunakan ekstraksi ciri *Fast Fourier Transform*

Bambang Sudibya<sup>1</sup>, Mardiana Irawaty<sup>2</sup>, Sajidah Nur Aizzun<sup>3</sup>, Sugiono<sup>4</sup>

<sup>1,2,3</sup>Program Studi Teknik Elektro Institut Teknologi Dirgantara Adisutjipto Yogyakarta

<sup>4</sup>Program Studi Teknik Elektro, Fakultas Universitas Islam Malang

## Article Info

### Article history:

Received January 25, 2024

Accepted February 21, 2024

Published March 1, 2024

### Kata Kunci :

Pengenalan suara,

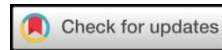
Sistem pengenalan emosi,

Matlab,

Transformasi forier cepat

## ABSTRAK

Pengenalan emosi ialah suatu proses untuk mengidentifikasi emosi pada manusia. Di mana emosi dapat dikenali melalui kata-kata, intonasi suara, ekspresi wajah, serta bahasa tubuh. Pada penelitian ini penulis merancang sistem pengenalan emosi dengan basis suara manusia menggunakan *software* matlab dengan ekstraksi ciri yang digunakan *fast fourier transform* yang merupakan proses ekstraksi ciri yang sederhana dengan mengubah sinyal dalam domain waktu diskrit menjadi domain waktu frekuensi. Pada penelitian ini berfokus pada empat kelas emosi: marah, netral, senang, serta sedih. Tahap pertama dimulai dari proses pengambilan data suara, pemotongan, *preprocessing*, proses ekstraksi *fast fourier transform* dan *dynamic time warping*. Dari hasil tersebut diperoleh nilai akurasi tertinggi pada emosi sedih sebesar 100% dan nilai akurasi lainnya diperoleh rata-rata sebesar 66,67%. Kemudian dari keseluruhan nilai akurasi yang diperoleh untuk seluruh kelas emosi, rata-rata nilai akurasinya adalah 75%.



## Corresponding Author:

Bambang Sudibya,

Program Studi Teknik Elektro, Institut Teknologi Dirgantara Adisutjipto Yogyakarta,

Jl. Raya Janti Blok R, Banguntapan, Kec. Banguntapan, Kabupaten Bantul, Daerah Istimewa Yogyakarta

Email: sudibya.stta@gmail.com

## 1. PENDAHULUAN

Ilmu pengetahuan dan teknologi pada generasi ini telah berkembang semakin pesat sehingga berdampak besar pada kehidupan manusia, salah satunya ialah sistem pengolahan sinyal suara. Pengolahan sinyal suara adalah perubahan sinyal suara yang disengaja dan terjadi melalui efek suara atau unit efek yang dipresentasikan secara elektronik dengan format analog maupun digital yang terjadi pada kedua domain. Pengaplikasian sinyal suara telah tersebar luas dikalangan masyarakat dalam berbagai bidang seperti musik, telekomunikasi, navigasi, medis, dan sebagainya [1].

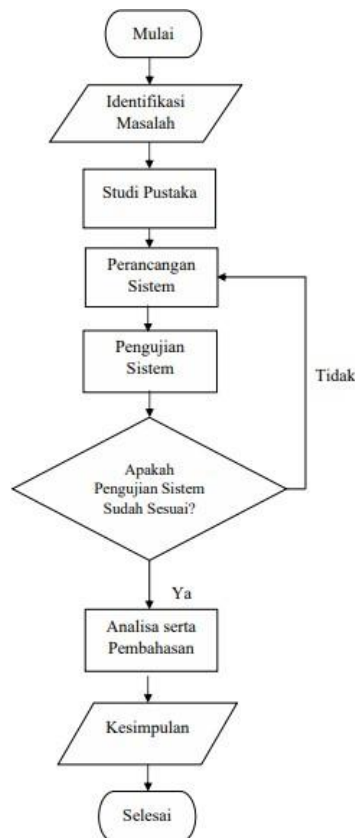
Suara sendiri ialah getaran yang dihasilkan oleh fenomena fisik berupa sinyal analog dengan amplitudo yang berubah secara *kontinyu* terhadap waktu dan merambat melalui media transmisi seperti cair, padat, dan gas [2]. Salah satu kegiatan yang menghasilkan suara ialah ketika manusia berbicara. Ketika manusia berbicara, manusia memiliki nada pengucapan yang berbeda – beda tergantung dengan suasana hati yang sedang dirasakan. Keadaan suasana hati disebut juga emosi. Untuk mengetahui keadaan emosi dapat dikenali melalui fitur *pitch* dan *formant*. Di mana *pitch* adalah getaran pita suara yang menghasilkan sinyal suara berupa frekuensi *fundamental*, sedangkan *formant* adalah bidang suara manusia yang menghasilkan frekuensi resonansi akustik [3].

Pada penelitian sebelumnya telah didapatkan tingkat keberhasilan 90% pada pengenalan emosi suara menggunakan metode *Hidden Markov Model* dengan ekstraksi fitur *Mel Frequency Cepstral Coefficient* [4]. Penelitian lainnya oleh Panca Mudji Raharjo, menggunakan algoritma jaringan nural network, hasil penelitian yaitu dapat mengenali wajah yang tinggi, namun disini hasilnya masih kualitatif dan penelitian ini masih perlu disempurnakan lagi [5]. Penelitian lainnya menggunakan beberapa ekstraksi wajah dan yang digunakan adalah

frekuensi dan hasilnya adalah sebesar 73,05% jadi penelitian ini masih dapat dimaksimalkan lagi hasilnya [6]. Dengan demikian, penulis tertarik untuk melakukan penelitian tentang pengenalan emosi menggunakan metode yang berbeda dari penelitian-penelitian sebelumnya yaitu dengan menggunakan *Fast Fourier Transform*, dimana *Fast Fourier Transform* (FFT) merupakan proses perubahan sinyal dalam domain waktu diskret menjadi domain waktu frekuensi dengan data [7] yang dihasilkan bersifat kuantitatif serta diharapkan memiliki akurasi pengenalan emosi yang lebih tinggi. Dengan proses FFT terhadap data latih dan data uji dapat dibedakan sehingga tidak terjadi kesalahan saat penggunaan data. Manfaat pengenalan emosi berbasis suara menggunakan FFT yaitu dapat diaplikasikan pada industri penerbangan dengan cara sistem tersebut digunakan untuk mendeteksi emosi pilot apakah pilot layak untuk menerbangkan pesawat atau tidak, jika pilot pada keadaan emosi dalam kategori mayor.

## 2. METODE PENELITIAN

Pada penelitian ini langkah pertama yang dilakukan adalah identifikasi masalah, kemudian dilakukan studi pustaka untuk mencari materi serta referensi yang berkaitan sebagai penunjang dasar teori dalam penelitian. Selanjutnya dilakukan perancangan sistem dengan dibantu perangkat keras maupun perangkat lunak. Untuk proses pengujian sistem dilakukan dengan memasukkan nilai data uji apakah memiliki nilai yang sama dengan data latih. Jika hasil yang didapatkan tidak sesuai maka proses perancangan sistem dilakukan ulang sampai didapatkan hasil yang sama. Jika memiliki hasil yang sama maka data tersebut diteruskan untuk dikaji lebih dalam agar menghasilkan kesimpulan dari penelitian yang dilakukan. Gambar 1 berikut menggambarkan diagram alir penelitian ini.



Gambar 1. Diagram Alir Penelitian

### 2.1 Pemotongan Suara

Sebelum melakukan pemotongan suara dilakukan proses perekaman suara dengan bantuan 8 responden untuk mengatakan kata “iya” sebanyak 5 kali dengan 4 kali proses perekaman sesuai emosi yang akan diteliti. Hasil dari perekaman suara akan diteruskan pada aplikasi *audacity* untuk melakukan pemotongan gelombang suara agar memiliki durasi selaras satu detik, lalu mengubah frekuensi suara agar menjadi 8000 Hz serta menyimpannya dengan format *.wav*.

## 2.2 Preprocessing

*Preprocessing* dilakukan guna memperbaiki kualitas objek yang mana memiliki tiga tahapan meliputi normalisasi, *frame blocking*, dan *windowing*.

### 2.2.1 Normalisasi

Tahap pertama pada *preprocessing* ialah normalisasi, dimana normalisasi digunakan untuk menyetarakan amplitudo suara yang terekam menjadi maksimum sehingga efek kuat atau lemahnya suara yang terekam tidak mempengaruhi proses pengenalan suara [8]. Normalisasi dilakukan dengan membagi tiap nilai data masukan suara terekam dengan nilai *absolute* maksimal dari data masukan. Nilai normalisasi dapat dicari dengan rumus Persamaan (1) sebagai berikut.

$$X_{norm} = \frac{x_{in}}{\max(abs(x_{in}))} \quad (1)$$

dengan:  $X_{norm}$  = hasil data sinyal

$x_{in}$  = data masukan dari sampling

$\max(abs(x_{in}))$  = nilai *absolute* maksimal sinyal masukan

### 2.2.2 Frame Blocking

Hasil dari proses normalisasi kemudian diteruskan pada proses *frame blocking*. Proses ini berguna untuk memudahkan perhitungan dan analisis suara dengan hanya mengenali suara yang dibutuhkan. *Frame N* diperoleh dari sinyal suara yang diblok. *Frame* pertama terdiri dari  $N$  sampel pertama. *Frame* kedua dengan  $M$  sampel setelah *frame* pertama, *overlap* dengan  $N-M$  sampel. Untuk mengetahui jumlah *frame* diperlukan Persamaan (2) yang dirumuskan sebagai berikut.

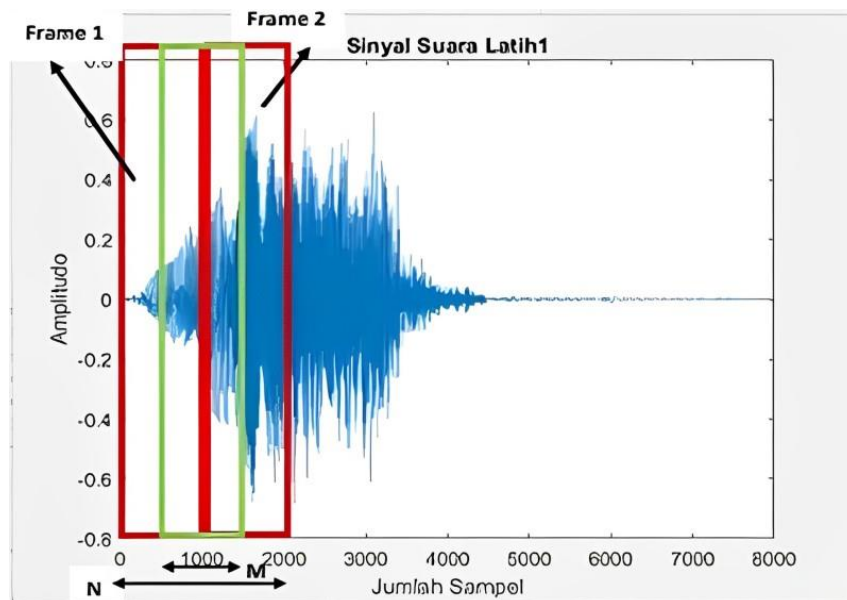
$$\text{Jumlah frame} = \frac{(I-N)}{M} + 1 \quad (2)$$

dengan:  $I$  = total sampel

$N$  = ukuran frame

$M$  = overlapping

Gambar 2 merupakan proses *frame blocking* yang dapat dilihat sebagai berikut.



Gambar 2. Proses *Frame Blocking*

### 2.2.3 Windowing

Tahap terakhir dari *preprocessing* adalah *windowing*. Proses *windowing* digunakan untuk mengurangi serta menghilangkan efek diskontinuitas pada ujung – ujung *frame*. Jenis *windowing* yang digunakan dalam proses penelitian ini ialah *windowing hamming*, karena memiliki *slide lobe* paling kecil dan *main lobe* yang paling besar sehingga akan mendapatkan hasil yang lebih halus dalam menghilangkan efek diskontinuitas [1].

### 2.3 Ekstraksi Ciri *Fast Fourier Transform*

Setelah melakukan *preprocessing*, proses selanjutnya ialah ekstraksi ciri dengan menggunakan *Fast Fourier Transform*. FFT dimulai dengan mencari nilai komputasi melalui tiga tahapan perhitungan yang selanjutnya mencari nilai *magnitude* [9]. Nilai *absolute* yang didapatkan dijadikan proses terakhir ekstraksi ciri. Berikut merupakan Persamaan (3) untuk ekstraksi ciri *fast fourier transform*.

$$s(f) = \int_{-\infty}^{\infty} s(t)e^{-j2\pi ft} dt \tag{3}$$

dengan :  $s(f)$  = sinyal dalam domain frekuensi

$(t)$  = sinyal dalam domain waktu

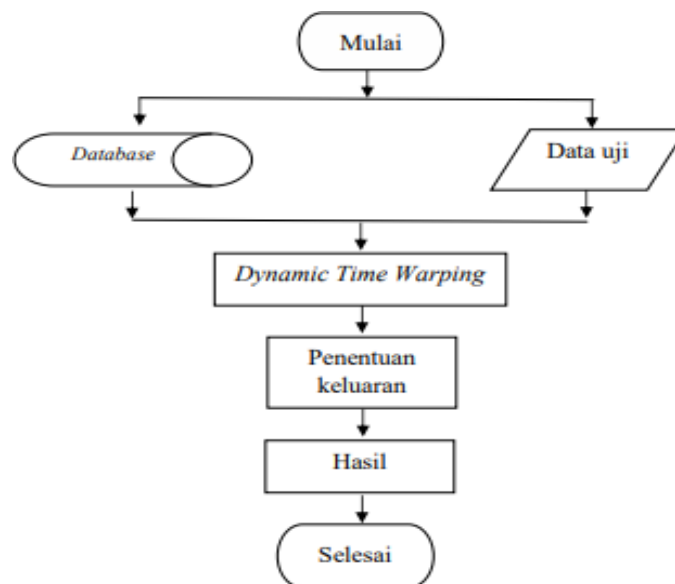
$\pi$  = derajat ( $180^\circ$ )

$e$  = konstanta dari nilai sebuah sinyal

$dt$  = waktu

### 2.4 Penentuan Keluaran dengan Klasifikasi *Dynamic Time Warping*

Pengklasifikasian pada penelitian ini untuk menentukan keluaran menggunakan *Dynamic Time Warping* dengan mengukur kemiripan suatu pola dengan membandingkan data latih dan data uji [10]. Hasil dari perhitungannya berupa selisih nilai terkecil yang mana jika semakin kecil jarak yang diperoleh maka kemiripan akan semakin besar begitupula sebaliknya yang kemudian ditentukan sebagai keluaran [11]. Proses pencocokan pola dengan metode *Dynamic Time Warping* dapat dilihat pada Gambar 3 berikut.



Gambar 3. Diagram Alir *Dynamic Time Warping*

## 3. HASIL DAN PEMBAHASAN

### 3.1 Proses Data Latih

Proses data latih dengan mengambil 5 data suara ditiap emosi, sehingga jumlah data yang dipakai sebanyak 20 data suara. Data suara yang telah didapatkan kemudian masuk pada proses pemotongan suara yang selanjutnya melewati perhitungan *preprocessing* dimana dalam proses ini terdapat tiga tahapan: normalisasi, *frame blocking*, dan *windowing*. Hasil dari *windowing* kemudian diteruskan untuk dilakukan proses ekstraksi ciri FFT. Hasil akhir yang didapatkan dari proses FFT itulah yang nantinya disebut data latih untuk dibandingkan dengan data uji.

**3.1.1 Normalisasi**

Hasil proses normalisasi untuk data latih dapat dilihat pada Tabel 1 berikut.

Tabel 1. Hasil Normalisasi untuk Data Latih

Data Suara	Hasil Normalisasi untuk Data Latih		
	1	2	.....8000
Latih1	8.977869551e-05	0.00089778696	-0.0003591148
Latih2	0.003417829678	-0.00911421248	0.000253172569
Latih3	0.003091986601	-0.00206132440	0.002641071889
Latih4	-0.014203070533	-0.04950678632	0
Latih5	0.003063164566	0.007781142173	-0.00239419759
Latih6	-0.026210063134	-0.05242012627	0.000382628659
Latih7	0.099606077659	0.192881260551	-0.00801913337
Latih8	0.034138626820	0.030321208486	-0.00076348367
Latih9	0.047348181381	-0.07979490466	-0.00184265342
Latih10	-0.019359001936	-0.03570660357	-0.00731340073
Latih11	-0.005722178374	-0.00749802683	-0.03670086819
Latih12	-0.010219467247	-0.01189478975	-0.01725582175
Latih13	0.018444457709	0.03491911896	0.011042798305
Latih14	-0.007990351274	-0.03241368913	-0.00331674959
Latih15	0.002730669733	0.01077895947	0.000574877838
Latih16	0.008332319669	0.01873251429	0.0029801727284
Latih17	-0.000198866461	0.00865069106	-0.0100427562891
Latih18	-0.010758611750	-0.00495803493	-0.005184873132
Latih19	0.015207534007	0.033275200558	0.0005929543076
Latih20	-0.014372542009	-0.012727922774	0.0018948873793

**3.1.2 Frame Blocking**

Setelah mendapatkan hasil dari normalisasi, maka proses selanjutnya ialah *frame blocking*. Pada penelitian ini panjang *frame* yang digunakan ialah 25 ms dengan *overlap* sebesar 50%, sehingga dari Persamaan (4) digunakan rumus sebagai berikut:

$$Jumlah\ frame = \frac{(1-N)}{M} + 1 \tag{4}$$

dengan :  $I = Fs \times durasi\ perekaman = 8000 \times 1 = 8000$

$N = I \times waktu\ framing = 8000 \times 0.25 = 200$

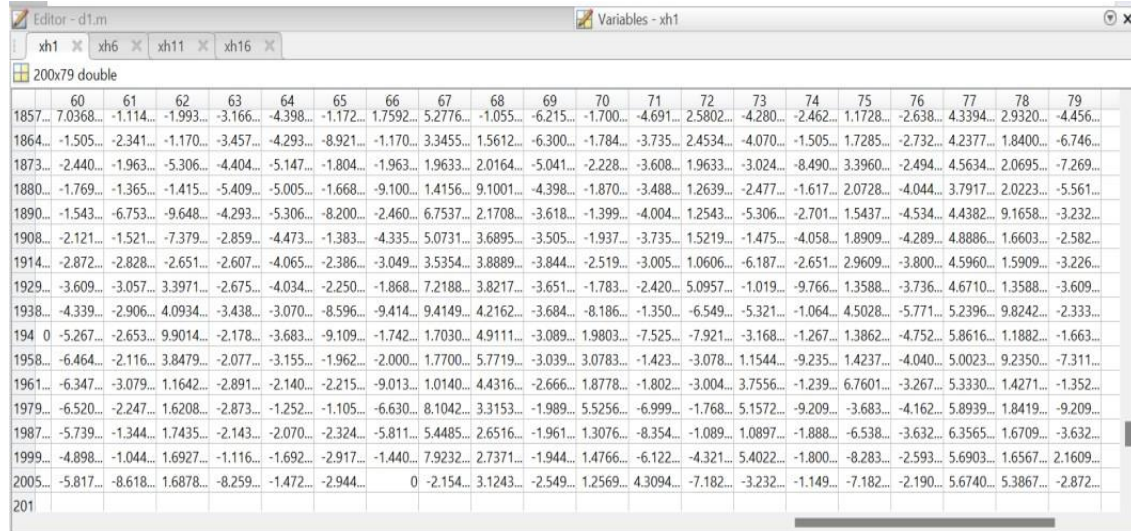
$M = 50\% \times N = 50\% \times 200 = 100$

Sehingga jumlah *frame* =  $\frac{(8000-200)}{100} + 1 = 79\ frame$

Dari perhitungan tersebut maka jumlah *frame* yang digunakan sebanyak 79 *frame* dimana terdapat 200 *sample* di tiap *framenya* dengan *overlapping* 100 *sample*, sehingga *frame* yang kedua dimulai dari data ke 101 hingga data ke 300. Begitu seterusnya hingga data ke 7800 sampai data ke 8000.

### 3.1.3 Windowing

Tahap selanjutnya ialah *windowing*. Pada penelitian ini jenis *windowing* yang digunakan adalah *windowing hamming* yang berfungsi meminimalisir *diskontinuitas* yang dapat menyebabkan kehilangan antar *frame*. Hasil perkalian sinyal suara dengan *windowing hamming* pada *workspace* matlab data latih 1 untuk emosi marah dapat dilihat pada Gambar 4, dimana untuk N adalah titik sampel 1,2,3,4,...200.



Gambar 4. Hasil *Windowing Hamming* Latih 1 (Marah)

### 3.1.4 Ekstraksi Ciri *Fast Fourier Transform*

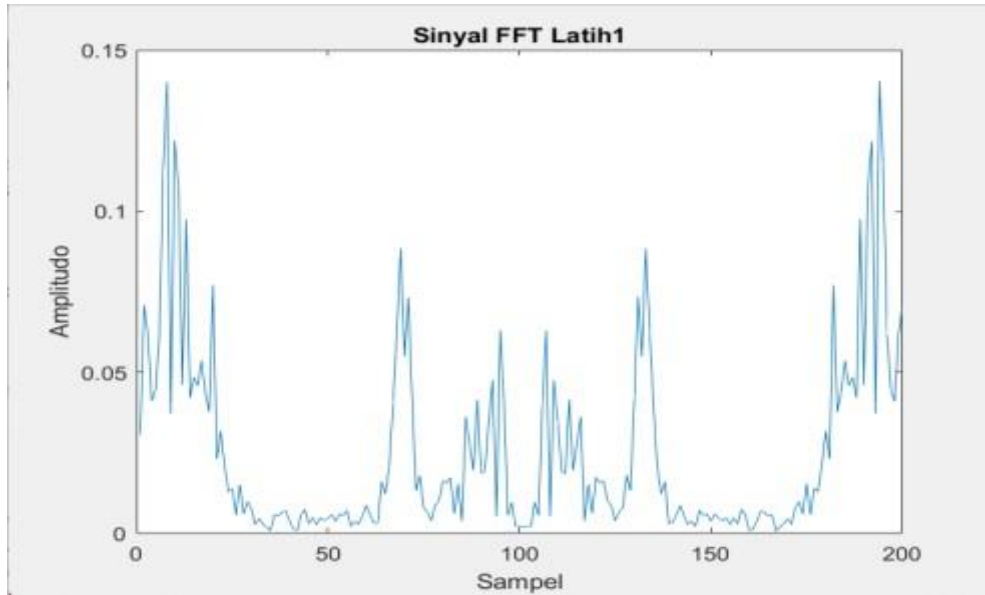
Hasil dari proses *Fast Fourier Transform* untuk data latih dapat dilihat pada Tabel 2 berikut.

Tabel 2. Hasil *Fast Fourier Transform* untuk Data Latih

Data Suara	Hasil <i>Fast Fourier Transform</i> untuk Data Latih		
	1	2	.....200
Latih1	0.030639979373	0.070643357015	0.0706433570145
Latih2	0.042346879643	0.040586447454	0.0405864474537
Latih3	0.086431729457	0.155955632407	0.1559556324067
Latih4	0.040761230258	0.026101116695	0.0261011166954
Latih5	0.001038299211	0.095784322264	0.0957843222610
Latih6	0.183161202298	0.439906321742	0.4399063217423
Latih7	0.403234799879	0.125314960893	0.1253149608927
Latih8	0.037939734683	0.282373339169	0.2823733391694
Latih9	0.027444428249	0.048782296028	0.0487822960278
Latih10	-0.019359001936	-0.035706603571	0.1287373628737
Latih11	-0.0057221783741	-0.007498026835	-0.3018942383583
Latih12	-0.0102194672474	-0.011894789747	-0.4841682023789
Latih13	0.0184444577091	0.0349191189638	-0.0629737957381
Latih14	-0.0079903512739	-0.032413689130	-0.0039197949646
Latih15	0.0027306697327	0.0107789594711	-0.0263006611095
Latih16	0.0083323196691	0.0187325142927	0.0853302517942
Latih17	-0.0001988664611	0.0086506910609	0.2814954757881
Latih18	-0.0107586117502	-0.0049580349331	-0.6960692180563
Latih19	0.0152075340077	0.0332752005581	-0.0588071154516
Latih20	-0.0143725420093	-0.0127279227744	0.05355738291026



Dari tabel tersebut dapat dilihat hasil FFT data latih 1, dimana untuk N adalah titik sampel 1,2,3,4,...200. Pada Gambar 5 merupakan grafik sinyal *fast fourier transform* dari data suara latih 1 untuk emosi marah yang diperlihatkan sebagai berikut.



Gambar 5. Sinyal FFT untuk Data Latih 1 (Marah)

Hasil dari FFT latih 1 untuk emosi marah pada *workspace* dapat dilihat pada Gambar 6 berikut, dimana untuk N adalah titik sampel 1,2,3,4,...200.

	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79
18670	0.0224	0.0120	0.0131	0.0095	0.0121	0.0020	0.0062	0.0102	0.0037	0.0114	0.0066	0.0038	0.0031	0.0049	0.0058	8.1153...	0.0062	0.0049	0.0098	0.0168
18742	0.0268	0.0103	0.0070	0.0118	0.0032	0.0119	0.0066	0.0098	0.0117	0.0023	0.0047	0.0110	0.0020	0.0025	0.0069	0.0099	0.0165	0.0052	0.0033	0.0141
18884	0.0283	0.0086	0.0128	0.0094	0.0039	0.0204	0.0109	0.0057	0.0106	0.0080	0.0195	0.0133	0.0053	0.0067	0.0065	0.0045	0.0128	0.0077	0.0078	0.0160
18920	0.0144	0.0098	0.0145	0.0077	0.0025	0.0147	0.0149	0.0052	0.0121	0.0122	0.0185	0.0089	0.0051	0.0074	0.0161	0.0072	0.0012	0.0135	0.0090	0.0102
19078	0.0643	0.0294	0.0047	0.0054	0.0036	0.0057	0.0125	0.0206	0.0138	0.0179	0.0068	0.0110	0.0076	0.0068	0.0190	0.0166	0.0137	0.0154	0.0082	0.0064
19109	0.0917	0.0023	0.0241	0.0321	0.0184	0.0157	0.0167	0.0138	0.0232	0.0172	0.0150	0.0084	0.0207	0.0027	0.0119	0.0040	0.0217	0.0228	0.0108	0.0250
19287	0.0498	0.0122	0.0144	0.0388	0.0185	0.0142	0.0110	0.0296	0.0074	0.0281	0.0067	0.0123	0.0215	0.0041	0.0197	0.0047	0.0179	0.0364	0.0111	0.0321
19332	0.1018	0.0253	0.0132	0.0180	0.0122	0.0174	0.0057	0.0142	0.0267	0.0250	0.0189	0.0076	0.0245	0.0320	0.0104	0.0302	0.0058	0.0279	0.0141	0.0310
19464	0.0686	0.0506	0.0335	0.0214	0.0246	0.0191	0.0115	0.0297	0.0133	0.0332	0.0114	0.0164	0.0220	0.0462	0.0414	0.0354	0.0276	0.0069	0.0214	0.0383
19552	0.0895	0.0321	0.0424	0.0296	0.0091	0.0370	0.0119	0.0248	0.0139	0.0296	0.0089	0.0148	0.0135	0.0120	0.0158	0.0076	0.0363	0.0105	0.0207	0.0176
19637	0.0365	0.0289	0.0785	0.0316	0.0024	0.0485	0.0249	0.0287	0.0578	0.0417	0.0286	0.0194	0.0191	0.0280	0.0157	0.0096	0.0373	0.0218	0.0166	0.0448
19743	0.2209	0.0611	0.0598	0.0548	0.0322	0.0794	0.0299	0.0340	0.0781	0.1063	0.0296	0.0173	0.0397	0.1218	0.0523	0.0395	0.0026	0.1175	0.0596	0.0411
19812	0.5489	0.2232	0.0808	0.1051	0.0556	0.0697	0.1425	0.0715	0.0459	0.1539	0.1447	0.1137	0.1171	0.1563	0.1776	0.0558	0.0536	0.2027	0.2320	0.0083
19936	0.5556	0.3680	0.0692	0.1625	0.1838	0.1182	0.1826	0.0677	0.1238	0.1189	0.1584	0.2499	0.1187	0.0555	0.2092	0.0662	0.1040	0.0889	0.2753	0.1558
20084	0.2597	0.3083	0.1265	0.0667	0.1534	0.0908	0.0650	0.0860	0.0975	0.0630	0.0678	0.1696	0.0724	0.0791	0.1018	0.0776	0.0446	0.0685	0.1439	0.1244

Gambar 6. Hasil FFT Latih 1 pada *Workspace* Matlab (Marah)

### 3.2 Proses Data Uji

Proses data uji dengan mengambil 3 data suara ditiap emosi, sehingga jumlah data yang dipakai sebanyak 12 data suara. Data suara yang telah didapatkan kemudian masuk pada proses pemotongan suara yang selanjutnya melewati perhitungan *preprocessing* dimana dalam proses ini terdapat tiga tahapan: normalisasi, *frame blocking*, dan *windowing*. Hasil dari *windowing* kemudian diteruskan untuk dilakukan proses ekstraksi ciri FFT. Hasil akhir yang didapatkan dari proses FFT itulah yang nantinya disebut data uji untuk dibandingkan dengan data latih.

### 3.2.1 Normalisasi dan Frame Blocking

Hasil proses normalisasi untuk data uji dapat dilihat pada Tabel 3 berikut.

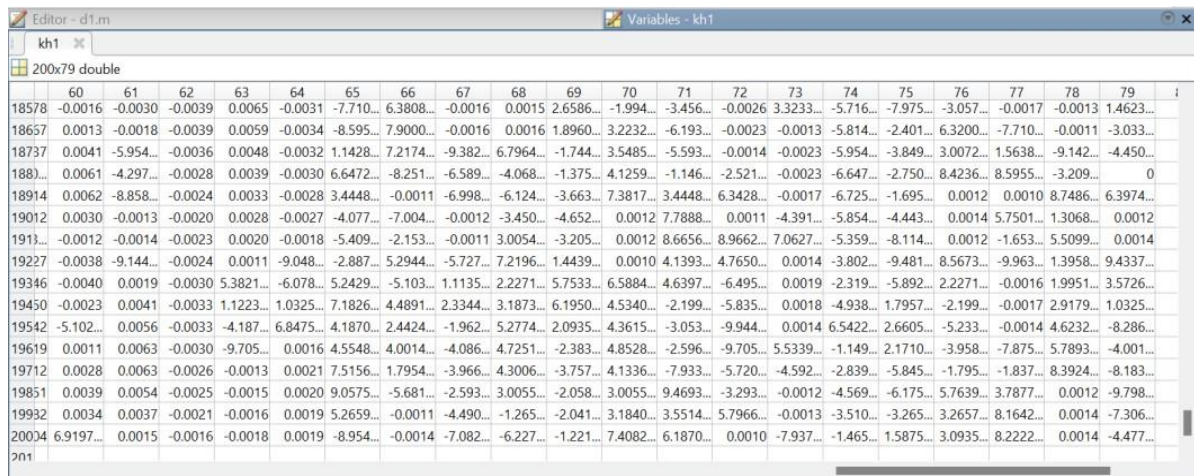
Tabel 3. Hasil Normalisasi untuk Data Uji

Data Suara	Hasil Normalisasi untuk Data Uji		
	1	2	.....8000
Uji1	-0.002136969573	-0.004680980971	-0.005596825074
Uji2	0.023298309730	-0.001862459149	-0.000773096250
Uji3	0.000853914871	-0.002955859169	-0.017767997898
Uji4	0.002421924793	0.003824091778	0
Uji5	-0.001306620209	-0.015080574913	-0.000272212544
Uji6	-0.000359130903	0.004219788113	-0.005297180822
Uji7	-0.002690582959	-0.003587443946	0.001614349776
Uji8	-0.002928109855	-0.007118336026	-0.000151453958
Uji9	0.0035149384886	0.024311657879	-0.025483304042
Uji10	-0.002452483139	-0.004775888218	-0.000613120785
Uji11	0.0042966983265	0.002002972152	0.000904568069
Uji12	-0.005730659026	-0.011059166541	-0.003619363595

Seperti dari perhitungan sebelumnya jumlah *frame* yang digunakan sebanyak 79 *frame* dimana terdapat 200 *sample* di tiap *framenya* dengan *overlapping* 100 *sample*, sehingga *frame* yang kedua dimulai dari data ke 101 hingga data ke 300. Begitu seterusnya hingga data ke 7800 sampai data ke 8000

### 3.2.2 Windowing

Tahap selanjutnya ialah *windowing*. Pada penelitian ini jenis *windowing* yang digunakan adalah *windowing hamming* yang berfungsi meminimalisir *diskontinuitas* yang dapat menyebabkan kehilangan antar *frame*. Hasil perkalian sinyal suara dengan *windowing hamming* pada *workspace* matlab data uji 1 dapat dilihat pada gambar berikut dimana untuk N adalah titik sampel 1,2,3,4,...200. Gambar 7 berikut merupakan hasil *Windowing Hamming Uji 1*.



Gambar. 7 Hasil *Windowing Hamming Uji 1* ( Merah )

### 3.2.3 Ekstraksi Ciri Fast Fourier Transform

Tabel 4. Hasil *Fast Fourier Transform* untuk Data Uji

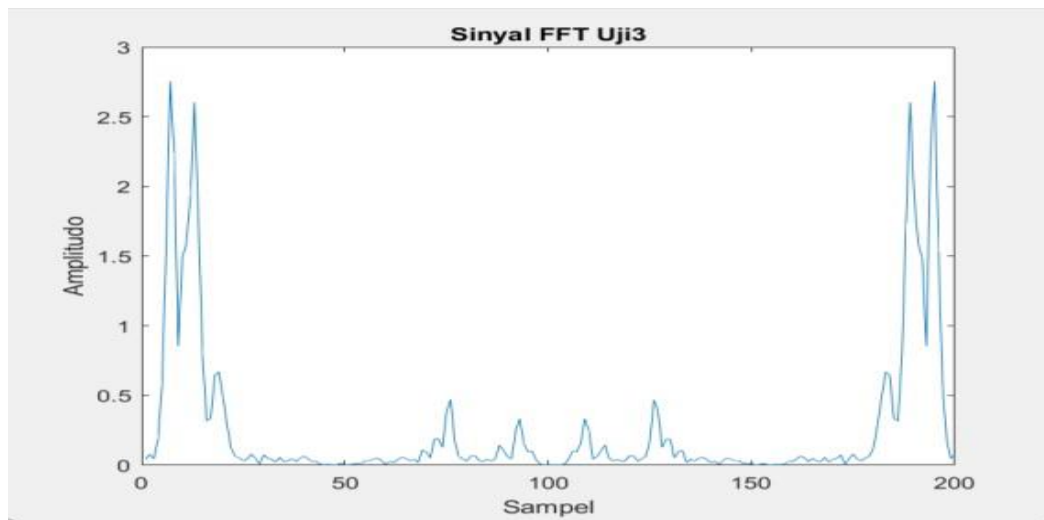
Data Suara	Hasil <i>Fast Fourier Transform</i> untuk Data Uji		
	1	2	.....200
Uji1	0.0225140578019	0.054538773446	0.0545387734435
Uji2	0.0184148117491	0.042253935286	0.0422539352857



Sistem pengenalan emosi berbasis suara menggunakan ekstraksi ciri *Fast Fourier Transform*

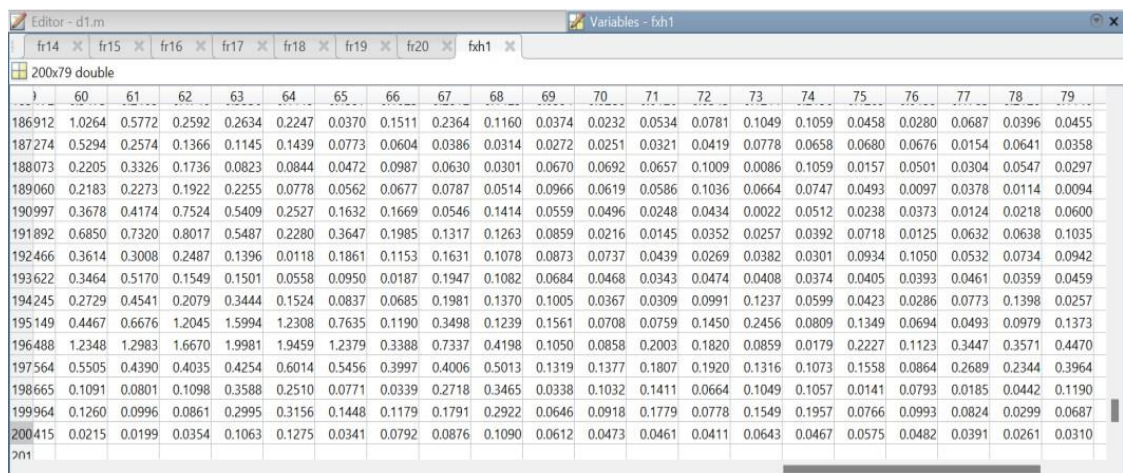
Uji3	0.0442807613221	0.077599616210	0.0775996162102
Uji4	0.1586498669447	0.159008347341	0.1590083473406
Uji5	-0.0001045296167	-0.001209903501	0.0166637630662
Uji6	-2.8730472226e-05	0.000338550516	0.0256203986353
Uji7	0.0038026905829	0.003122825358	0.0004161434978
Uji8	-0.0002342487883	-0.000571098896	0.0193295638126
Uji9	0.000281195079	0.001950506539	0.014364381956
Uji10	-0.000196198651	-0.000383166021	-0.000379489496
Uji11	0.0003437358661	0.000160696992	-0.010526587840
Uji12	-0.000458452722	-0.000887268847	0.004463881767

Dari Tabel 4 dapat dilihat hasil FFT data uji 3, dimana untuk N adalah titik sampel 1,2,3,4,...200. Berikut merupakan grafik sinyal *fast fourier transform* dari data suara uji 3 untuk emosi marah yang diperlihatkan Gambar 8 berikut.



Gambar 8. Sinyal FFT Uji 3 (Marah)

Hasil dari FFT uji 1 untuk emosi marah pada *workspace* dapat dilihat pada Gambar 9 sebagai berikut, dimana untuk N adalah titik sampel 1,2,3,4,...200.



Gambar 9. Hasil FFT Uji 1 pada *Workspace* Matlab (Marah)

### 3.3 Proses Pengujian Sistem

Nilai hasil proses *Fast Fourier Transform* yang diperoleh dari pengujian data latih dan data uji kemudian dieksekusi menggunakan klasifikasi *Dynamic Time Warping*. Hasil pengujian berwarna hijau

menunjukkan jika suara dapat dikenali sedangkan pengujian yang berwarna hijau menunjukkan jika suara tersebut tidak dikenali dengan kata lain *error*. Pengujian untuk emosi marah dapat dilihat pada Tabel 5 berikut.

Tabel 5. Hasil Uji untuk Emosi Marah

	Uji 1	Uji 2	Uji 3
Latih 1	1.8945e+03	1.3141e+03	2.9262e+03
Latih 2	2.2014e+03	1.9424e+03	3.4907e+03
Latih 3	2.4112e+03	1.9336e+03	3.2889e+03
Latih 4	1.6198e+03	1.6336e+03	2.8108e+03
Latih 5	1.8560e+03	1.5680e+03	3.1110e+03
Latih 6	1.7192e+03	1.8888e+03	3.0988e+03
Latih 7	2.5788e+03	2.5201e+03	3.4407e+03
Latih 8	2.1437e+03	1.9465e+03	3.2114e+03
Latih 9	1.6988e+03	1.7709e+03	2.9687e+03
Latih 10	1.9557e+03	1.9146e+03	3.0736e+03
Latih 11	1.6869e+03	1.7917e+03	3.1021e+03
Latih 12	2.2064e+03	2.3158e+03	3.1752e+03
Latih 13	2.4115e+03	2.4584e+03	2.2825e+03
Latih 14	1.9926e+03	1.8978e+03	3.2465e+03
Latih 15	2.0468e+03	1.9298e+03	3.1482e+03
Latih 16	2.3935e+03	2.1729e+03	2.5623e+03
Latih 17	2.7638e+03	2.6780e+03	2.3875e+03
Latih 18	2.8552e+03	3.1091e+03	4.1497e+03
Latih 19	1.8113e+03	1.8443e+03	2.8606e+03
Latih 20	1.8437e+03	1.6447e+03	3.1052e+03

Pengujian untuk emosi netral dengan klasifikasi *dynamic time warping* dapat dilihat pada Tabel 6 berikut.

Tabel 6. Hasil Uji untuk Emosi Netral

	Uji 4	Uji 5	Uji 6
Latih 1	1.4673e+03	1.5427e+03	1.3001e+03
Latih 2	2.2321e+03	2.4430e+03	2.1176e+03
Latih 3	2.0956e+03	1.9587e+03	2.0640e+03
Latih 4	837.5609	1.1983e+03	921.7834
Latih 5	1.3350e+03	1.2671e+03	1.2743e+03
Latih 6	931.5768	1.3460e+03	1.1160e+03
Latih 7	1.6024e+03	1.4097e+03	1.7519e+03
Latih 8	1.2526e+03	37.9881	1.2341e+03
Latih 9	1.1670e+03	1.4964e+03	690.5782
Latih 10	1.1545e+03	1.3370e+03	936.5900
Latih 11	797.4692	1.1639e+03	728.8840
Latih 12	1.3738e+03	1.2567e+03	1.4056e+03
Latih 13	2.2821e+03	2.2674e+03	2.1503e+03
Latih 14	1.3681e+03	1.7391e+03	854.8166
Latih 15	1.2046e+03	1.4873e+03	1.0331e+03
Latih 16	2.1281e+03	2.0251e+03	2.0486e+03
Latih 17	2.4558e+03	2.1850e+03	2.2172e+03
Latih 18	3.2871e+03	2.9930e+03	3.1540e+03
Latih 19	1.3621e+03	1.6075e+03	1.3534e+03
Latih 20	1.3631e+03	1.3252e+03	1.2152e+03

Pengujian untuk emosi sedih dengan klasifikasi *dynamic time warping* dapat dilihat pada Tabel 7 berikut.

Tabel 7. Hasil Uji untuk Emosi Sedih

	Uji 7	Uji 8	Uji 9
Latih 1	1.9396e+03	1.7989e+03	3.0238e+03
Latih 2	2.6493e+03	2.4977e+03	3.4818e+03
Latih 3	2.5531e+03	2.3609e+03	3.4984e+03
Latih 4	1.4909e+03	1.4082e+03	2.1441e+03
Latih 5	1.6975e+03	1.5653e+03	2.7971e+03
Latih 6	1.6957e+03	1.4654e+03	2.1824e+03
Latih 7	1.3125e+03	1.7886e+03	1.6471e+03
Latih 8	1.3701e+03	1.5531e+03	2.1453e+03
Latih 9	1.7311e+03	1.4598e+03	2.5684e+03
Latih 10	1.6112e+03	1.4928e+03	2.2354e+03
Latih 11	1.4047e+03	1.3205e+03	1.8066e+03
Latih 12	1.2751e+03	1.4690e+03	1.2282e+03
Latih 13	2.4369e+03	2.3652e+03	3.0170e+03
Latih 14	1.8149e+03	1.4738e+03	2.6099e+03
Latih 15	1.7085e+03	1.5240e+03	2.7267e+03
Latih 16	2.4251e+03	2.3183e+03	3.1670e+03
Latih 17	2.3323e+03	2.3841e+03	3.0714e+03
Latih 18	3.4883e+03	3.3555e+03	4.1293e+03
Latih 19	1.9452e+03	1.7924e+03	2.7849e+03
Latih 20	1.6674e+03	1.6152e+03	2.8966e+03

Pengujian untuk emosi senang dengan klasifikasi *dynamic time warping* dapat dilihat pada Tabel 8 berikut.

Tabel 8. Hasil Uji untuk Emosi Senang

	Uji 10	Uji 11	Uji 12
Latih 1	2.4984e+03	1.3203e+03	2.5358e+03
Latih 2	3.0566e+03	2.1740e+03	2.9526e+03
Latih 3	2.1680e+03	2.0717e+03	2.7038e+03
Latih 4	2.2727e+03	1.2902e+03	2.8530e+03
Latih 5	2.2729e+03	1.5237e+03	2.5277e+03
Latih 6	2.5606e+03	1.5649e+03	3.0524e+03
Latih 7	3.0930e+03	2.2481e+03	3.5389e+03
Latih 8	2.1193e+03	1.7420e+03	3.0226e+03
Latih 9	2.4890e+03	1.3323e+03	2.9365e+03
Latih 10	2.5886e+03	1.5676e+03	2.8201e+03
Latih 11	2.5205e+03	1.4803e+03	2.9695e+03
Latih 12	2.8413e+03	1.9967e+03	3.2354e+03
Latih 13	2.8618e+03	1.5971e+03	3.1650e+03
Latih 14	2.6556e+03	1.6881e+03	3.0303e+03
Latih 15	2.5892e+03	1.6526e+03	2.8998e+03
Latih 16	2.5647e+03	1.8653e+03	3.0104e+03
Latih 17	2.7316e+03	1.7848e+03	3.1162e+03
Latih 18	2.3103e+03	3.1013e+03	3.2719e+03
Latih 19	2.2979e+03	1.3824e+03	3.0868e+03
Latih 20	2.0493e+03	1.3476e+03	2.4725e+03

### 3.4 Analisis Hasil

Dari hasil yang diperoleh dalam proses percobaan maka dilakukan analisis guna menetapkan dan mengambil keputusan dalam mengatasi suatu masalah agar mendapatkan kesimpulan yang akurat. Pengklasifikasian yang digunakan dalam penelitian ini ialah *Dynamic Time Warping*. Hasil dari pengujian terdiri dari empat bagian untuk emosi marah, netral, sedih, dan senang. Sehingga didapatkan analisis sebagai berikut:

#### 3.4.1 Analisis Hasil Pengujian *Dynamic Time Warping* untuk kelas emosi “marah”

Proses klasifikasi dengan metode *Dynamic Time Warping* untuk kelas emosi marah terdiri dari 5 data latih dengan nama *file* latih1, latih2, latih3, latih4 dan latih5 serta 3 data uji dengan nama *file* uji1, uji2, dan uji3. Dari 5 data latih tersebut kemudian dibandingkan dengan data lainnya yang berjumlah 20 guna mencari nilai minimum data yang diuji untuk mengetahui keakuratannya. Hasil dari DTW kelas emosi marah dapat dilihat pada Tabel 5. Dari tabel tersebut dapat ditarik kesimpulan jika nilai pada data uji1 dan uji2 *output* terkecil terdapat pada area data latih1 – latih5 yang berarti data tersebut *valid*. Uji1 nilai minimum terdapat pada latih4 sebesar  $1.6198e+03$  dan uji2 nilai minimum terdapat pada latih1 sebesar  $1.3141e+03$ . Sedangkan pada uji3 diartikan sebagai data *error* dikarenakan nilai minimum tidak terdapat pada area latih kelas emosi marah melainkan terdapat pada data latih13 sebesar  $2.2825e+03$ , dimana latih13 merupakan area untuk kelas emosi sedih.

#### 3.4.2 Analisis Hasil Pengujian *Dynamic Time Warping* untuk kelas emosi “netral”

Proses klasifikasi dengan metode *Dynamic Time Warping* untuk kelas emosi netral terdiri dari 5 data latih dengan nama *file* latih6, latih7, latih8, latih9 dan latih10 serta 3 data uji dengan nama *file* uji4, uji5, dan uji6. Dari 5 data latih tersebut kemudian dibandingkan dengan data lainnya yang berjumlah 20 guna mencari nilai minimum data yang diuji untuk mengetahui keakuratannya. Hasil dari DTW kelas emosi netral dapat dilihat pada Tabel 6. Dari tabel tersebut dapat ditarik kesimpulan jika nilai pada data uji8 dan uji9 *output* terkecil terdapat pada area data latih6 – latih1 yang berarti data tersebut *valid*. Uji8 nilai minimum terdapat pada latih8 sebesar 37.9881 dan uji9 nilai minimum terdapat pada latih9 sebesar 690.5782. Sedangkan pada uji6 diartikan sebagai data *error* dikarenakan nilai minimum tidak terdapat pada area latih kelas emosi sedih melainkan terdapat pada data latih11 sebesar 797.4692 dimana latih11 merupakan area untuk kelas emosi sedih.

#### 3.4.3 Analisis Hasil Pengujian *Dynamic Time Warping* untuk kelas emosi “sedih”

Proses klasifikasi dengan metode *Dynamic Time Warping* untuk kelas emosi sedih terdiri dari 5 data latih dengan nama *file* latih11, latih12, latih13, latih14 dan latih15 serta 3 data uji dengan nama *file* uji7, uji8, dan uji9. Dari 5 data latih tersebut kemudian dibandingkan dengan data lainnya yang berjumlah 20 guna mencari nilai minimum data yang diuji untuk mengetahui keakuratannya. Hasil dari DTW kelas emosi sedih dapat dilihat pada Tabel 7. Dari tabel tersebut dapat ditarik kesimpulan jika nilai minimum ke-tiga data uji yang diujikan terdapat pada area data latih antara latih11 – latih15 yang menunjukkan jika data tersebut *valid*. Uji7 dengan nilai minimum  $1.2751e+03$  pada latih12, uji8 dengan nilai minimum  $1.3205e+03$  pada latih11, dan uji9 dengan nilai minimum  $1.2282e+03$  pada latih12. Sehingga pada pengujian kelas emosi sedih tidak terjadi *error*.

#### 3.4.4 Analisis Hasil Pengujian *Dynamic Time Warping* untuk kelas emosi “senang”

Proses klasifikasi dengan metode *Dynamic Time Warping* untuk kelas emosi senang terdiri dari 5 data latih dengan nama *file* latih16, latih17, latih18, latih19 dan latih20 serta 3 data uji dengan nama *file* uji10, uji11, dan uji12. Dari 5 data latih tersebut kemudian dibandingkan dengan data lainnya yang berjumlah 20 guna mencari nilai minimum data yang diuji untuk mengetahui keakuratannya. Hasil dari DTW kelas senang dapat dilihat pada Tabel 8. Dari tabel tersebut dapat ditarik kesimpulan jika nilai pada data uji10 dan uji12 *output* terkecil terdapat pada area data latih16 – latih20 yang berarti data tersebut *valid*. Uji10 nilai minimum terdapat pada latih20 sebesar  $2.0493e+03$  dan uji12 nilai minimum terdapat pada latih2 sebesar  $2.4725e+03$ . Sedangkan pada uji11 diartikan sebagai data *error* dikarenakan nilai minimum tidak terdapat pada area latih kelas emosi senang melainkan terdapat pada data latih4 sebesar  $1.2902e+03$ , dimana latih4 merupakan area untuk kelas emosi marah.

### 3.5 Akurasi Pengenalan Emosi dengan Suara

Persamaan (5) yang digunakan untuk menghitung akurasi pengenalan emosi dengan suara sebagai berikut [12]:

$$\% \text{ Akurasi} = \frac{\sum \text{Data yang dikenali}}{\sum \text{Data uji}} \times 100\% \quad (5)$$

$$\% \text{ Akurasi emosi marah} = \frac{2}{3} \times 100\% = 66.67\%$$

$$\% \text{ Akurasi emosi netral} = \frac{2}{3} \times 100\% = 66.67\%$$

$$\% \text{ Akurasi emosi sedih} = \frac{3}{3} \times 100\% = 100\%$$

$$\% \text{ Akurasi emosi senang} = \frac{2}{3} \times 100\% = 66.67\%$$

$$\% \text{ Akurasi rata - rata} = \frac{66.67+66.67+100+66.67}{4} \times 100\% = 75\%$$

Dari perhitungan tersebut didapatkan hasil nilai akurasi untuk tiap emosi dan hasil rata-rata akurasi untuk seluruh emosi seperti pada Tabel 9 berikut.

Tabel 9. Hasil Akurasi

No	Emosi	Jumlah datayang diuji untuk tiap emosi	Jumlah datayang dikenali	Hasil Akurasi
1.	Marah	3	2	66.67%
2.	Netral	3	2	66.67%
3.	Sedih	3	3	100%
4.	Senang	3	2	66.67%
Rata-rata pengenalan				75%

Dari Tabel 9 dapat disimpulkan jika akurasi untuk kelas emosi marah sebesar 66.67%, akurasi kelas emosi netral sebesar 66.67%, akurasi kelas emosi sedih 100% dan akurasi untuk kelas emosi senang sebesar 66.67%. Sehingga nilai rata-rata pengenalan emosi menggunakan suara dengan membagi jumlah hasil akurasi seluruh emosi dengan jumlah kelas emosi jika dibulatkan sebesar 75%.

#### 4. KESIMPULAN

Berdasarkan proses pengerjaan selama perancangan, implementasi, dan proses pengujian dari penelitian ini, maka dapat diambil beberapa kesimpulan. Pertama, merancang sistem untuk mengenali emosi berdasarkan data suara dengan membandingkan hasil dari data latih dan data uji yang didapatkan dari hasil ekstraksi ciri *fast fourier transform* yang sebelumnya terdapat duatahapan berupa pengambilan data suara dan *preprocessing*. Kedua, Sistem yang telah dirancang menghasilkan nilai rata-rata akurasi sebesar 75% dengan nilai akurasi tertinggi sebesar 100% untuk pengenalan kelas emosi sedih dan nilai akurasi Sebesar 66.67% untuk pengenalan kelas emosi lainnya meliputi marah, netral, serta senang. Yang terakhir yaitu proses pengklasifikasian menggunakan *Dynamic Time Warping* dengan mengukur kemiripan suatu pola dengan mencari nilai minimum atau terkecil.

Penelitian lebih lanjut diperlukan untuk memperbaiki nilai akurasi lebih sempurna dengan memastikan data suara memiliki *noise* sekecil mungkin agar dapat dikenali oleh sistem dengan jelas. Pengembangan berikutnya diharapkan dapat mengenali emosi dengan data lain, bisa dengan gambar wajah tidak hanya dengan suara saja. Selain itu, dapat pula dikembangkan kriteria kelayakan terbang berdasarkan emosi pilot yang di deteksi dengan FFP terkait aplikasinya dalam industri penerbangan.

#### DAFTAR PUSTAKA

- [1] A. Riyani. "A Identifying Human Voice Signals Using the Fast Fourier Transform (Fft) Method Based on Matlab," *Journal of Informatics Information System Software Engineering and Applications (INISTA)*, vol.1, no. 2, hal. 42-50, 2019. <https://doi.org/10.20895/inista.v1i2.52>
- [2] L. S. Kristina. "Pemisahan Suara Manusia Berdasarkan Jenis Kelamin Menggunakan *Fast Fourier Transform* (FFT)," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 7, no. 3, hal. 610-616, 2020. <https://doi.org/10.35957/jatisi.v7i3.461>
- [3] C. Busso, S. Lee, S. Narayanan. "Analysis of emotionally salient aspects of fundamental frequency for emotion detection," *IEEE transactions on audio, speech, and language processing*, vol. 17, no. 4, hal. 582-596, 2009. <https://doi.org/10.1109/TASL.2008.2009578>
- [4] S. Helmiyah, A. Fadlil, A. Yudhana. "Pengenalan Pola Emosi Manusia Berdasarkan Ucapan Menggunakan Ekstraksi Fitur Mel-Frequency Cepstral Coefficients (MFCC)," *CogITO Smart Journal*, vol. 4, no. 2, hal. 372-381, 2019. <https://doi.org/10.31154/cogito.v4i2.129.372-381>
- [5] Panca M. Raharjo. "Pengenalan Ekspresi Wajah Berbasis Filter Gabor Dan Backpropagation Neural Network." *EECCIS*, vol. 4, no. 1, 2010. <https://doi.org/10.21776/jeccis.v4i1.96>
- [6] L. D. P. Bagaswari. "Implementasi algoritma backpropagation pada pengenalan emosi berdasarkan suara manusia", Fakultas Sains dan Teknologi, Universitas Islam Negeri Syarif Hidayatullah, Jakarta.



- 2019.
- [7] I. Mohidin dan F. Tupamahu. "Analisis Koefisien Cepstral Emosi Berdasarkan Suara," *Journal of Applied Informatics and Computing (JAIC)*, vol. 1, no. 1, 2017. <https://doi.org/10.30871/jaic.v1i2.523>
- [8] H. Kurniawati. "Media Audio Visual Sebagai Sarana Pengenalan Ekspresi Emosi," *KOMUNIKA: Jurnal Dakwah dan Komunikasi*, vol. 6, no.2, 2012. <https://doi.org/10.24090/komunika.v6i2.353>
- [9] R. Umar, I. Riadi, A. Hanif, "Analisis Bentuk Pola Suara Menggunakan Ekstraksi Ciri Mel Frequency Cepstral Coefficients (MFCC)," *CogITO Smart J.*, vol. 4, no. 2, hal. 294, 2019. <https://doi.org/10.31154/cogito.v4i2.130.294-304>.
- [10] H. Cho. "Comparing English and Korean speakers' word-final/r/ clusters using dynamic time warping," *Phonetics and Speech Sciences*, vol. 14, no. 1, hal. 29-36, 2022. <https://doi.org/10.13064/KSSS.2022.14.1.029>
- [11] Friedrich Alexander. "Article Dynamic Time Warping" *Sensors*, vol. 15, no. 3, hal. 6419-6440, 2007. <https://doi.org/10.3390/s150306419>
- [12] B. H. Prasetyo, W. Kurniawan, M. H. H. Ichsan. "Pengenalan emosi berdasarkan suara menggunakan algoritma hmm," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)* vol. 4, hal. 168-172, 2017. <https://doi.org/10.25126/jtik.201743339>