

BAB IV

HASIL DAN PEMBAHASAN

4.1. Pengujian Perangkat Lunak

Dalam mengetahui perangkat lunak yang dibuat bisa sesuai dengan metode yang dipakai maka dilakukan pengujian terhadap masing-masing komponen perangkat. Alat dan bahan yang digunakan sebelum pengujian:

1. *Personal Computer (PC) / Laptop.*
2. *Smartphone berbasis Android.*
3. *Software Adobe Audition 2014.*
4. *Software Matlab 2013.*
5. *Microsoft Excel 2016.*
6. *Aplikasi Android Smart Voice Recorder.*

4.2. Proses Rekaman dan Edit File Rekaman

Setelah suara direkam menggunakan *smartphone*, maka file hasil rekaman tersebut diedit atau dipotong secara manual menggunakan perangkat lunak *Adobe Audition* untuk menghilangkan rekaman kosong.

Tabel 4.1. – tabel 4.10. menunjukkan waktu file rekaman pada laki-laki dan perempuan pada saat selesai perekaman dan setelah diedit (crop). Contohnya pada perekaman huruf “a” pada responden laki-laki pertama, sinyal suara manusia baru terbentuk saat 0.653 s dan berakhir pada 1.480 s, sehingga sinyal suara yang diambil adalah sinyal pada rentang 0.653 s dan 1.480 s, seperti ditunjukkan pada tabel 4.1. Dengan demikian waktu keseluruhan sinyal suara adalah 0.827 s seperti

terlihat pada tabel 4.1. begitu juga pada responden laki-laki kedua dan huruf vokal lainnya (e, i, o, u):

Tabel 4.1. File rekaman sinyal suara “a” laki-laki

<i>Nama File</i>	<i>Waktu keseluruhan rekaman (s)</i>	<i>Waktu awal crop (s)</i>	<i>Waktu akhir crop (s)</i>	<i>Waktu rekaman setelah crop (s)</i>
<i>a_1</i>	2.414	0.653	1.480	0.827
<i>a_2</i>	1.393	0.337	0.900	0.563
<i>a_3</i>	1.300	0.315	1.002	0.686
<i>a_4</i>	1.578	0.460	1.198	0.738
<i>a_5</i>	1.027	0.426	0.880	0.453
<i>a_6</i>	1.114	0.052	0.799	0.747
<i>a_7</i>	2.043	0.457	1.159	0.702
<i>a_8</i>	1.393	0.375	0.969	0.594
<i>a_9</i>	1.857	0.448	1.020	0.571
<i>a_10</i>	1.393	0.100	0.830	0.729
<i>a_11</i>	1.021	0.026	0.830	0.803
<i>a_12</i>	1.486	0.000	0.879	0.879
<i>a_13</i>	1.764	0.374	1.327	0.953
<i>a_14</i>	2.414	1.506	1.919	0.413
<i>a_15</i>	2.136	0.503	1.469	0.966

Tabel 4.2. File rekaman sinyal suara “e” laki-laki

<i>Nama File</i>	<i>Waktu keseluruhan rekaman (s)</i>	<i>Waktu awal crop (s)</i>	<i>Waktu akhir crop (s)</i>	<i>Waktu rekaman setelah crop (s)</i>
<i>e_1</i>	2.322	0.633	1.320	0.686
<i>e_2</i>	1.300	0.348	0.999	0.651
<i>e_3</i>	1.300	0.390	1.013	0.621
<i>e_4</i>	1.300	0.219	1.002	0.783
<i>e_5</i>	1.114	0.173	0.800	0.627
<i>e_6</i>	1.671	0.260	1.099	0.839
<i>e_7</i>	1.671	0.339	1.199	0.859
<i>e_8</i>	1.671	0.494	1.240	0.745

<i>e_9</i>	1.857	0.371	0.879	0.507
<i>e_10</i>	1.207	0.168	0.880	0.711
<i>e_11</i>	1.114	0.000	0.880	0.880
<i>e_12</i>	2.229	0.066	1.060	0.993
<i>e_13</i>	1.486	0.138	0.839	0.701
<i>e_14</i>	1.114	0.390	0.839	0.448
<i>e_15</i>	2.136	0.368	1.150	0.782

Tabel 4.3. File rekaman sinyal suara “i” laki-laki

<i>Nama File</i>	<i>Waktu keseluruhan rekaman (s)</i>	<i>Waktu awal crop (s)</i>	<i>Waktu akhir crop (s)</i>	<i>Waktu rekaman setelah crop (s)</i>
<i>i_1</i>	2.322	0.658	1.499	0.840
<i>i_2</i>	1.393	0.322	0.950	0.625
<i>i_3</i>	1.486	0.547	1.198	0.650
<i>i_4</i>	1.671	0.409	1.103	0.694
<i>i_5</i>	1.207	0.379	0.916	0.536
<i>i_6</i>	1.486	0.302	1.287	0.984
<i>i_7</i>	1.857	0.636	1.539	0.903
<i>i_8</i>	1.578	0.415	1.160	0.745
<i>i_9</i>	1.857	0.538	0.999	0.461
<i>i_10</i>	1.486	0.420	1.160	0.740
<i>i_11</i>	1.950	0.210	1.076	0.865
<i>i_12</i>	1.300	0.000	0.800	0.800
<i>i_13</i>	2.136	0.466	1.331	0.865
<i>i_14</i>	1.207	0.435	0.819	0.384
<i>i_15</i>	1.207	0.229	0.879	0.650

Tabel 4.4. File rekaman sinyal suara “o” laki-laki

<i>Nama File</i>	<i>Waktu keseluruhan rekaman (s)</i>	<i>Waktu awal crop (s)</i>	<i>Waktu akhir crop (s)</i>	<i>Waktu rekaman setelah crop (s)</i>
<i>o_1</i>	2.972	0.793	1.699	0.905
<i>o_2</i>	1.486	0.408	1.050	0.642
<i>o_3</i>	1.857	0.405	1.094	0.688

<i>o_4</i>	1.300	0.373	0.997	0.624
<i>o_5</i>	1.486	0.448	1.104	0.566
<i>o_6</i>	1.764	0.280	1.270	0.990
<i>o_7</i>	1.486	0.392	1.140	0.748
<i>o_8</i>	1.671	0.566	1.229	0.663
<i>o_9</i>	1.486	0.228	0.799	0.570
<i>o_10</i>	1.578	0.515	1.129	0.614
<i>o_11</i>	1.764	0.435	1.420	0.984
<i>o_12</i>	1.950	0.526	1.500	0.974
<i>o_13</i>	1.764	0.090	1.059	0.969
<i>o_14</i>	2.136	0.388	0.850	0.462
<i>o_15</i>	2.229	0.585	1.380	0.795

Tabel 4.5. File rekaman sinyal suara “u” laki-laki

<i>Nama File</i>	<i>Waktu keseluruhan rekaman (s)</i>	<i>Waktu awal crop (s)</i>	<i>Waktu akhir crop (s)</i>	<i>Waktu rekaman setelah crop (s)</i>
<i>u_1</i>	1.950	0.603	1.370	0.766
<i>u_2</i>	1.393	0.343	0.950	0.606
<i>u_3</i>	1.671	0.347	1.047	0.687
<i>u_4</i>	2.414	0.526	1.242	0.714
<i>u_5</i>	2.043	0.477	1.025	0.728
<i>u_6</i>	1.578	0.309	1.240	0.930
<i>u_7</i>	2.507	0.550	1.380	0.829
<i>u_8</i>	1.486	0.223	0.863	0.639
<i>u_9</i>	1.857	0.692	1.229	0.537
<i>u_10</i>	1.671	0.476	1.139	0.663
<i>u_11</i>	1.764	0.250	1.216	0.966
<i>u_12</i>	2.136	0.520	1.500	0.979
<i>u_13</i>	1.764	0.335	1.101	0.766
<i>u_14</i>	1.950	0.587	0.999	0.412
<i>u_15</i>	1.300	0.143	0.949	0.805

Dapat dilihat pada tabel 4.1. – 4.5. bahwa waktu *file* rekaman suara laki-laki setelah dipotong tidak boleh melebihi 1 detik. Begitu juga pada *file* rekaman suara perempuan seperti terdapat pada tabel 4.6. – tabel 4.10.

Berikut adalah tabel waktu rekaman pada perempuan:

Tabel 4.6. File rekaman sinyal suara “a” perempuan

<i>Nama File</i>	<i>Waktu keseluruhan rekaman (s)</i>	<i>Waktu awal crop (s)</i>	<i>Waktu akhir crop (s)</i>	<i>Waktu rekaman setelah crop (s)</i>
<i>a_1</i>	1.393	0.425	0.960	0.534
<i>a_2</i>	1.857	0.535	1.402	0.867
<i>a_3</i>	1.300	0.284	0.950	0.665
<i>a_4</i>	1.207	0.000	0.990	0.990
<i>a_5</i>	1.440	0.841	1.400	0.558
<i>a_6</i>	1.520	0.320	1.300	0.980
<i>a_7</i>	1.520	0.341	0.855	0.513
<i>a_8</i>	1.440	0.199	0.900	0.700
<i>a_9</i>	1.840	0.850	1.840	0.982
<i>a_10</i>	2.080	1.200	2.080	0.880
<i>a_11</i>	1.157	0.206	0.620	0.413
<i>a_12</i>	1.400	0.858	1.360	0.502
<i>a_13</i>	1.782	0.145	1.110	0.965
<i>a_14</i>	1.400	0.000	0.880	0.880
<i>a_15</i>	1.353	0.070	0.790	0.720

Tabel 4.7. File rekaman sinyal suara “e” perempuan

<i>Nama File</i>	<i>Waktu keseluruhan rekaman (s)</i>	<i>Waktu awal crop (s)</i>	<i>Waktu akhir crop (s)</i>	<i>Waktu rekaman setelah crop (s)</i>
<i>e_1</i>	1.486	0.498	1.000	0.500
<i>e_2</i>	1.857	0.380	1.299	0.918
<i>e_3</i>	1.207	0.344	0.880	0.536
<i>e_4</i>	1.300	0.145	1.139	0.994
<i>e_5</i>	1.440	0.479	1.079	0.600

<i>e_6</i>	1.600	0.630	1.600	0.959
<i>e_7</i>	1.600	0.472	1.240	0.767
<i>e_8</i>	1.520	0.445	1.270	0.775
<i>e_9</i>	2.240	1.200	2.160	0.960
<i>e_10</i>	2.400	1.400	2.260	0.860
<i>e_11</i>	1.286	0.302	0.720	0.417
<i>e_12</i>	1.020	0.425	0.880	0.455
<i>e_13</i>	1.518	0.370	1.190	0.820
<i>e_14</i>	2.107	0.400	1.250	0.850
<i>e_15</i>	1.132	0.171	0.840	0.668

Tabel 4.8. File rekaman sinyal suara “i” perempuan

<i>Nama File</i>	<i>Waktu keseluruhan rekaman (s)</i>	<i>Waktu awal crop (s)</i>	<i>Waktu akhir crop (s)</i>	<i>Waktu rekaman setelah crop (s)</i>
<i>i_1</i>	1.486	0.515	1.050	0.535
<i>i_2</i>	2.322	0.539	1.469	0.930
<i>i_3</i>	2.414	0.483	1.098	0.615
<i>i_4</i>	1.486	0.379	1.249	0.870
<i>i_5</i>	1.440	0.479	1.079	0.600
<i>i_6</i>	1.600	0.630	1.600	0.969
<i>i_7</i>	1.520	0.420	1.000	0.580
<i>i_8</i>	1.200	0.245	0.922	0.677
<i>i_9</i>	2.080	1.200	2.080	0.880
<i>i_10</i>	2.160	1.000	1.940	0.940
<i>i_11</i>	1.400	0.550	0.943	0.393
<i>i_12</i>	1.240	0.292	0.860	0.568
<i>i_13</i>	1.092	0.228	0.900	0.671
<i>i_14</i>	1.200	0.000	0.780	0.780
<i>i_15</i>	1.268	0.220	0.940	0.720

Tabel 4.9. File rekaman sinyal suara “o” perempuan

<i>Nama File</i>	<i>Waktu keseluruhan rekaman (s)</i>	<i>Waktu awal crop (s)</i>	<i>Waktu akhir crop (s)</i>	<i>Waktu rekaman setelah crop (s)</i>
<i>o_1</i>	1.393	0.443	0.950	0.507
<i>o_2</i>	1.578	0.143	1.136	0.993
<i>o_3</i>	1.114	0.315	0.876	0.561
<i>o_4</i>	1.578	0.598	1.578	0.980
<i>o_5</i>	1.200	0.320	0.819	0.499
<i>o_6</i>	1.680	0.582	1.470	0.887
<i>o_7</i>	1.760	0.665	1.400	0.734
<i>o_8</i>	1.440	0.289	1.020	0.730
<i>o_9</i>	1.680	0.800	1.680	0.880
<i>o_10</i>	2.560	1.476	2.380	0.903
<i>o_11</i>	1.289	0.476	0.830	0.353
<i>o_12</i>	1.104	0.230	0.720	0.490
<i>o_13</i>	1.953	0.589	1.370	0.780
<i>o_14</i>	1.836	0.600	1.480	0.880
<i>o_15</i>	1.378	0.315	0.950	0.635

Tabel 4.10. File rekaman sinyal suara “u” perempuan.

<i>Nama File</i>	<i>Waktu keseluruhan rekaman (s)</i>	<i>Waktu awal crop (s)</i>	<i>Waktu akhir crop (s)</i>	<i>Waktu rekaman setelah crop (s)</i>
<i>u_1</i>	1.300	0.273	0.850	0.576
<i>u_2</i>	2.043	0.670	1.662	0.992
<i>u_3</i>	1.486	0.306	0.951	0.644
<i>u_4</i>	2.043	0.735	1.720	0.984
<i>u_5</i>	1.520	0.710	1.400	0.690
<i>u_6</i>	1.840	0.620	1.600	0.980
<i>u_7</i>	1.680	0.463	1.100	0.646
<i>u_8</i>	1.520	0.456	1.080	0.623
<i>u_9</i>	2.000	1.200	2.000	0.800
<i>u_10</i>	2.240	1.200	2.050	0.850
<i>u_11</i>	1.163	0.325	0.771	0.446

u_{12}	1.398	0.590	1.190	0.600
u_{13}	1.201	0.145	0.899	0.754
u_{14}	2.000	0.516	1.510	0.993
u_{15}	1.471	0.276	0.900	0.623

4.3. Ekstraksi Ciri *Short Time Fourier Transform*

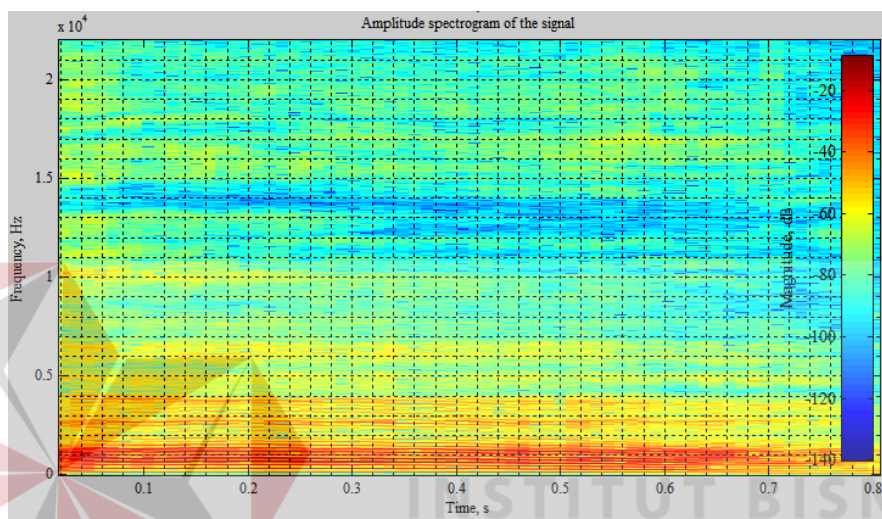
Setelah dipotong, maka *file* suara tersebut akan diekstraksi fiturnya menggunakan metode *short time fourier transform* (STFT). Nilai-nilai yang diekstraksi pada penelitian ini adalah nilai STFT, nilai rata-rata amplitudo pada sebuah sinyal, nilai frekuensi dan nilai waktu saat amplitudo tertinggi. Keempat nilai tersebut digunakan karena metode STFT menghasilkan suatu nilai waktu dan frekuensi, dan juga amplitudo.

Keempat nilai tersebut digunakan karena metode STFT menghasilkan suatu nilai waktu dan frekuensi, dan juga amplitude. Nilai STFT adalah suatu bilangan kompleks yang terdiri dari bilangan *real* dan *imaginer* yang merepresentasikan suatu nilai dalam domain waktu dan frekuensi. Agar nilai kompleks tersebut dapat dihitung, maka nilai tersebut diabsolutkan dengan rumus $absolut = \sqrt{real^2 + imaginer^2}$. Dan nilai amplitudo/magnitude adalah nilai yang menunjukkan kekuatan sinyal hasil keluaran dari mikropon yaitu pengukur atau penunjuk tingkat kebisingan suara atau *sound pressure level*, biasanya dalam satuan desibel (dB).

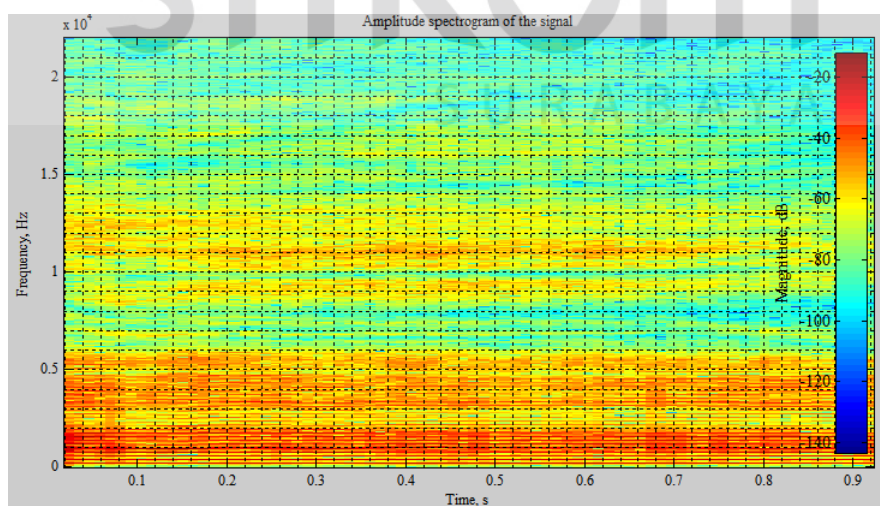
Kemudian program akan menampilkan grafik dan spektrogram pada setiap file yang diekstraksi. Banyaknya sinyal suara responden yang akan diekstraksi adalah 30 suara. Terdiri dari suara laki-laki saat mengucapkan a, e, i, o, u, masing-

masing berjumlah 15, dan suara perempuan saat mengucapkan a, e, i, o, u, masing-masing berjumlah 15.

Gambar 4.1. adalah spektrogram sinyal suara yang diambil dari tiga orang responden laki-laki pada saat mengucapkan huruf a, e, i, o, u, dengan menggunakan fungsi *windows hann* dengan panjang *windows* 2048.



Gambar 4.1. Spektrogram sinyal suara salah satu responden laki-laki saat pengucapan huruf “a”



Gambar 4.2. Spektrogram sinyal suara salah satu responden perempuan saat pengucapan huruf “a”

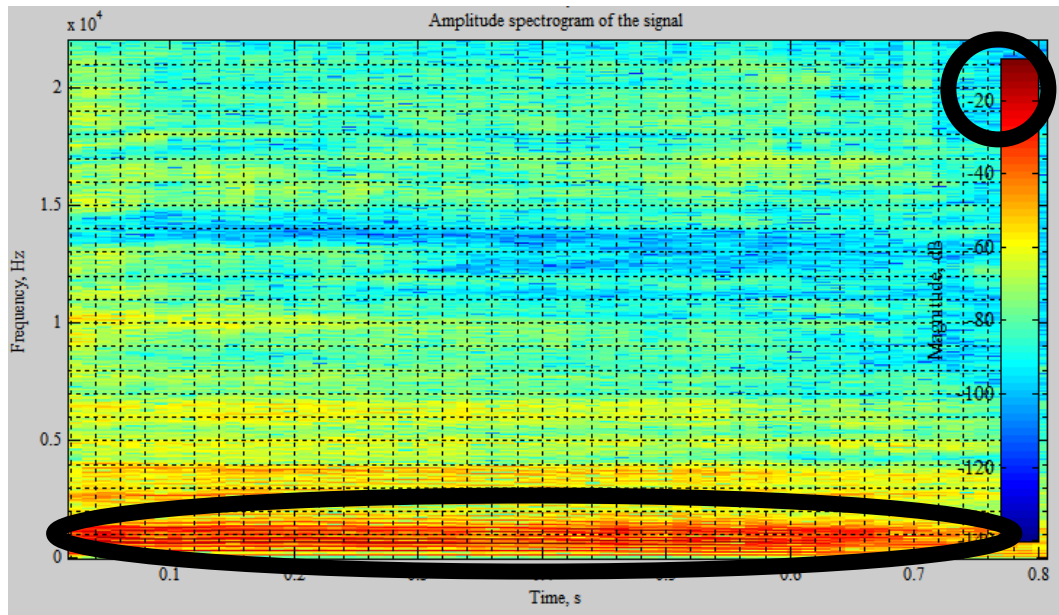
Gambar 4.1 menunjukkan representasi sinyal dalam domain waktu pada sumbu x dan domain frekuensi pada sumbu y. Sedangkan untuk kekuatan sinyal atau magnitude/amplitudo direpresentasikan dalam domain warna dimana warna merah adalah amplitudo tertinggi dan warna biru adalah amplitudo yang rendah. Adapun amplitudo adalah suatu matriks yang berisi nilai – nilai amplitudo dalam domain frekuensi dan waktu. Seperti contoh pada tabel 4.11.

Tabel 4.11. Contoh Matriks amplitudo sinyal suara salah satu responden laki-laki saat pengucapan huruf “a” dalam desibel (dB)

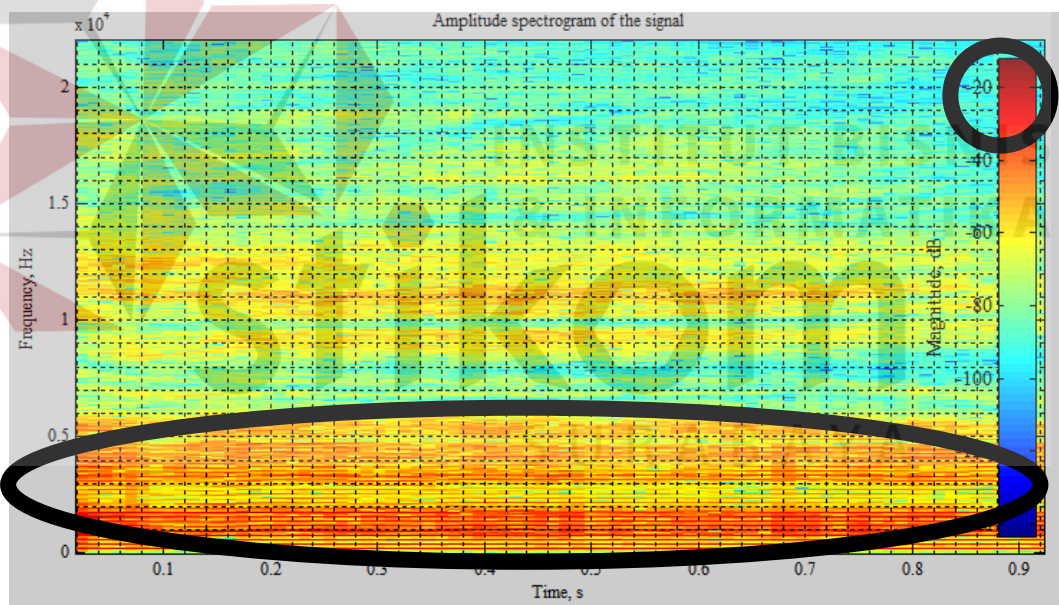
		Waktu (s)				
		0.0232	0.0348	0.0464	0.8011
Frekuensi (Hz)	10.7666	-66.7921	-71.8604	-75.7484	-72.0541
	21.5332	-61.2602	-67.1518	-70.3838	-67.6143
	32.2998	-63.1970	-73.0488	-74.0664	-71.2844
	43.0664	-66.4689	-88.0166	-88.1871	-75.2654

	22050	-95.6715	-122.0789	-89.3424	-108.6666

Matriks amplitudo pada tabel 4.11. adalah matriks yang berukuran besar. Dimana pada kolom adalah dalam domain waktu, dan baris adalah dalam domain frekuensi. Pada gambar 4.2., warna yang menunjukkan amplitudo tertinggi pada sinyal suara laki-laki tersebar di beberapa daerah.



Gambar 4.3. Letak amplitudo tertinggi pada spektrogram sinyal suara laki-laki



Gambar 4.4. Letak amplitudo tertinggi pada spektrogram sinyal suara perempuan

Gambar 4.2. menunjukkan lokasi amplitudo tinggi pada spektrogram sinyal suara laki-laki, dan gambar 4.3 menunjukkan lokasi amplitudo tertinggi pada spektrogram sinyal suara perempuan, dimana untuk spektrogram perempuan *range* frekuensi untuk amplitudo tertinggi lebih besar. Untuk mengetahui dimana letak amplitudo yang paling tinggi sangatlah sulit karena dalam spektrum warna,

beberapa warna terlihat sama, sehingga cara untuk mengetahui letaknya adalah dengan menganalisa matriks amplitudo pada tabel 4.11.

Pada aplikasi *Matlab*, menentukan letak elemen tertinggi pada suatu matriks adalah dengan sintaks sebagai berikut.

```

dBSA(:);
[SA_value,indexSA_no] = max(dBSA(:));
[SAI_row, SAI_col] = ind2sub(size(dBSA),indexSA_no);
sa_value = max(max(dBSA));

```

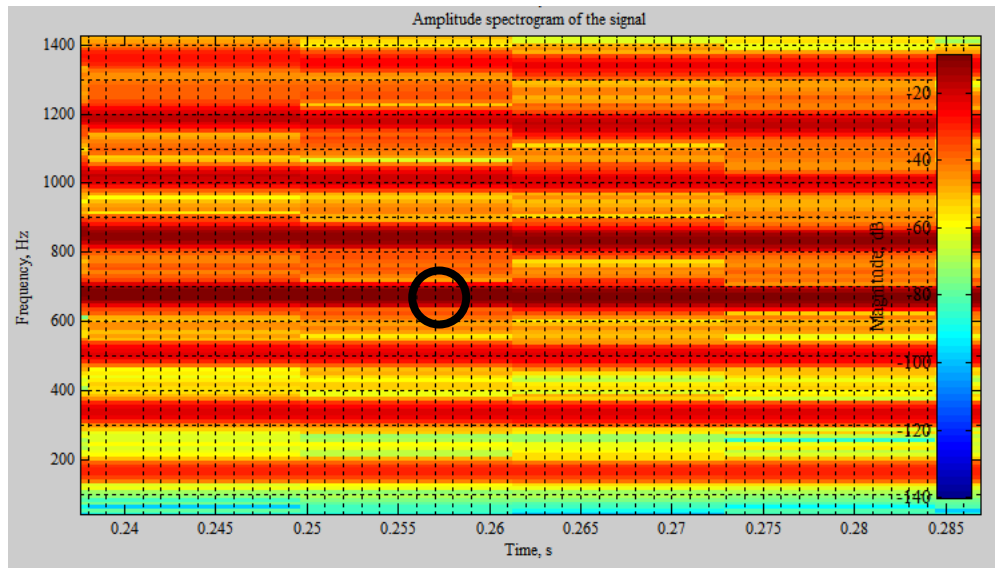
dBSA adalah matriks amplitudo yang berukuran (i x n), dimana dalam *Matlab*, semua elemen dapat diakses dengan sintaks `dBSA(:);`. Untuk mencari letak amplitudo tertinggi berada di elemen keberapa, maka digunakan sintaks

```

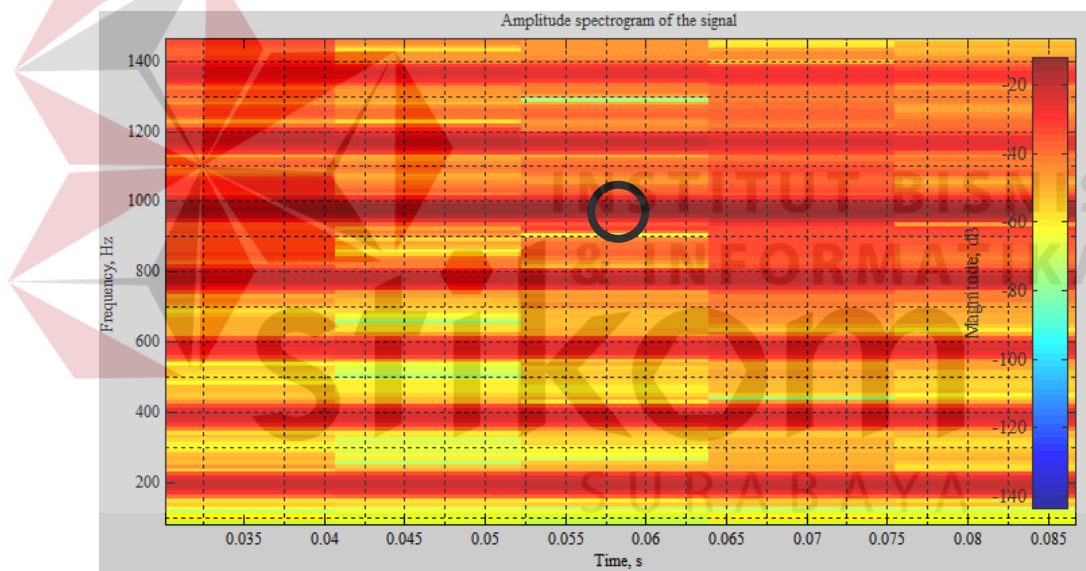
[SA_value,indexSA_no] = max(dBSA(:));
[SAI_row, SAI_col] = ind2sub(size(dBSA),indexSA_no);
sa_value = max(max(dBSA));

```

SAI_row dan **SAI_col** adalah lokasi amplitudo tertinggi, dan **sa_value** adalah nilai amplitudo tertinggi. Sehingga nilai yang didapatkan adalah -8.4734 dB. Waktu saat amplitudo tertinggi adalah 0.2554 s dan frekuensi saat amplitudo tertinggi adalah 678.2959 Hz, seperti teletak pada gambar 4.5. Sedangkan untuk perempuan, amplitudo tertinggi adalah -11.9230 dB. Waktu saat amplitudo tertinggi adalah 0.0580 s dan frekuensi saat amplitudo tertinggi adalah 979.7607 Hz, seperti terletak pada gambar 4.6.



Gambar 4.5. Letak amplitudo tertinggi pada sinyal suara laki-laki



Gambar 4.6. Letak amplitudo tertinggi pada sinyal suara perempuan

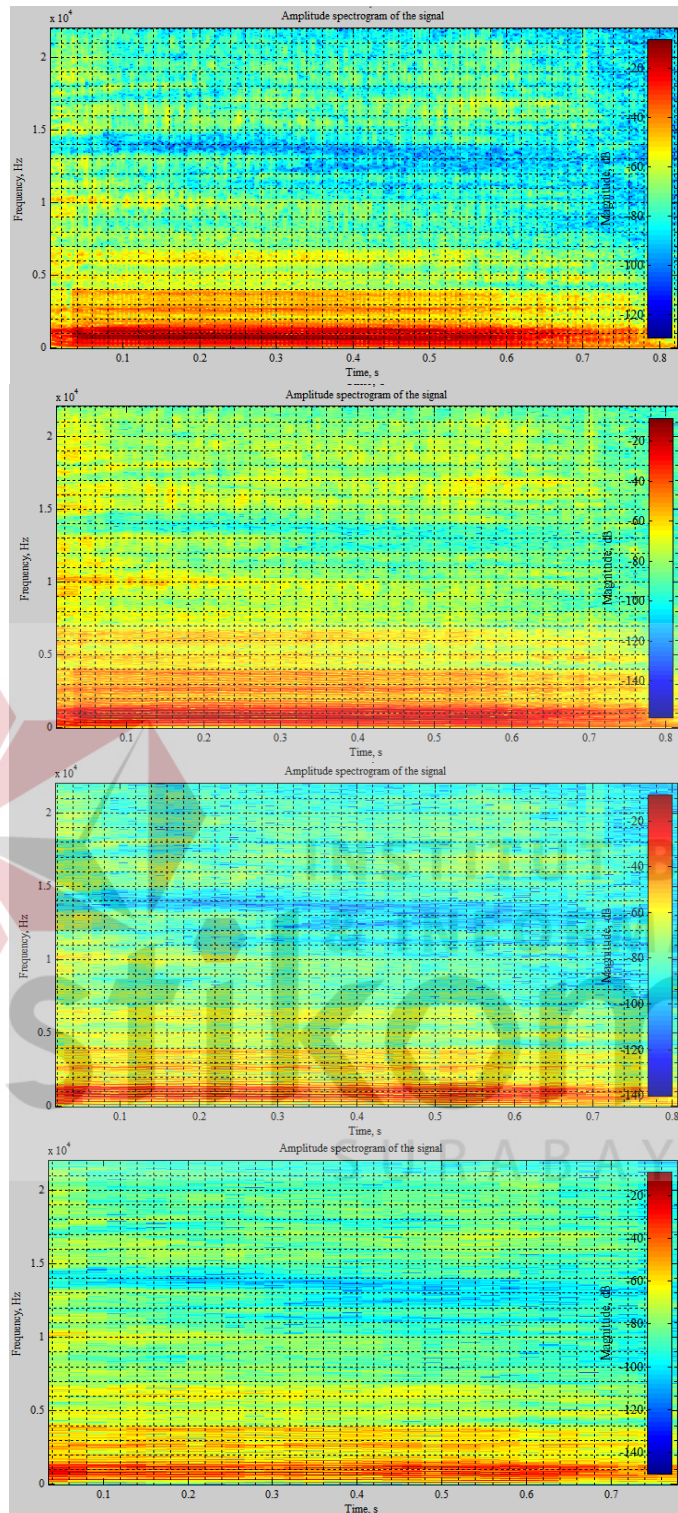
Gambar 4.5. dan gambar 4.6. adalah perbesaran dari gambar 4.4. dan gambar 4.5. sehingga amplitudo tertinggi untuk suara laki-laki saat pengucapan huruf “a” dapat diketahui terletak saat waktu 0.2554 s dan saat frekuensi 678.2959 Hz seperti teletak pada lingkaran hitam dalam gambar 4.5. dan amplitudo tertinggi untuk suara perempuan saat pengucapan huruf “a” dapat diketahui terletak saat

waktu 0.0580 s dan saat frekuensi 979.7607 Hz seperti teletak pada lingkaran hitam dalam gambar 4.6.

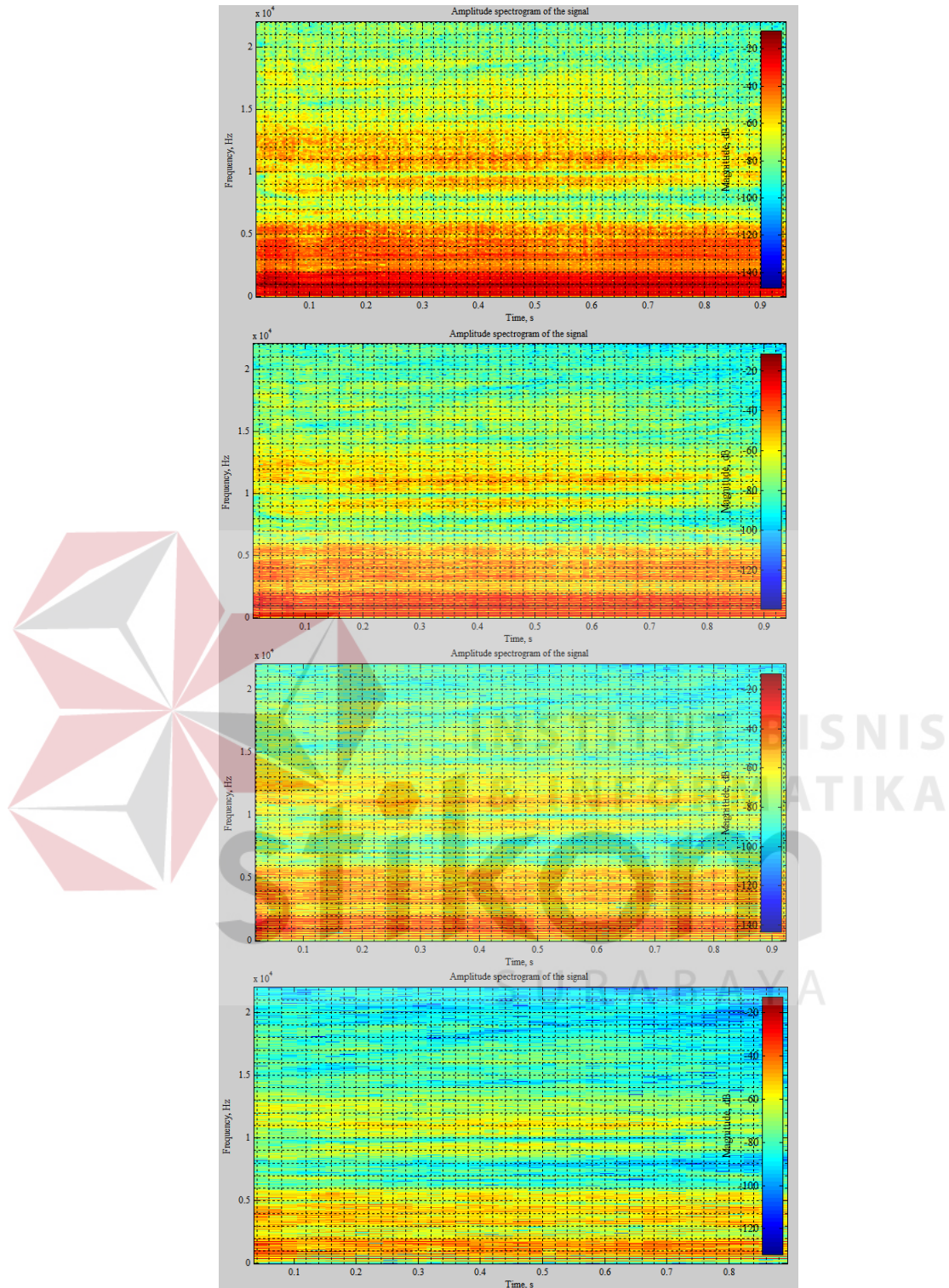
Nilai STFT adalah suatu bilangan kompleks hasil perhitungan transformasi *fourier* yang dikalikan dengan fungsi *window*, yang merepresentasikan suatu nilai dalam domain waktu dan frekuensi. Pada *Matlab*, hasil perhitungan dapat dicari dengan sintaks $X = \text{fft}(xw, \text{nfft});$

dimana $xw = x(\text{indx}+1:\text{indx}+\text{wlen}).*\text{win};$

xw adalah perhitungan nilai sampel pada sinyal suara sebanyak panjang *window* yang digunakan dan kemudian dikalikan dengan fungsi *window* (*win*). dan fungsi *window* yang digunakan adalah *hann window*. Panjang *window* yang digunakan adalah 2048. Nilai tersebut disesuaikan dengan *range* frekuensi dan waktu yang digunakan pada penelitian ini. Jika menggunakan ukuran *windows* yang lebih kecil, maka akan mempengaruhi *range* dari frekuensi dan waktu. Dan *nfft* adalah *fft points* dimana didapatkan dengan rumus $2*\text{panjang windows}$.



Gambar 4.7. Spektrogram sinyal suara salah satu responden laki-laki saat pengucapan huruf “a” dengan panjang *window* 512, 1024, 2048, dan 4096.



Gambar 4.8. Spektrogram sinyal suara salah satu responden perempuan saat pengucapan huruf “a” dengan panjang *window* 512, 1024, 2048, dan 4096

Gambar 4.7. dan gambar 4.8. menunjukkan panjang *window* yang berbeda beda. Perbedaan pada tiap spektrogram tersebut adalah jumlah matriks yang merepresentasikan waktu dan frekuensi dalam kolom dan baris. Seperti tercantum dalam tabel 4.12.

Tabel 4.12. ukuran matriks pada panjang *window* yang berbeda

Panjang <i>window</i>	Ukuran Matriks
512	513xn
1024	1025xn
2048	2049xn
4096	4096xn

Karena jangkauan *range* frekuensi pada penelitian ini sangat besar yaitu dari 0 Hz sampai 22050 Hz, namun jangkauan *range* waktu yang kecil yaitu antara 0 s – 1 s, maka, digunakanlah panjang *window* 2048 pada tabel 4.14. untuk mengoptimalkan jangkauan *range* frekuensi dan waktu. 2049 adalah *range* untuk frekuensi dan n adalah *range* untuk waktu. Adapun nilai n bervariasi tergantung dengan panjangnya waktu pada data rekaman.

Nilai STFT adalah suatu nilai berupa bilangan kompleks seperti terlihat pada tabel 4.13.

Tabel 4.13. Contoh sebagian nilai STFT dalam bilangan kompleks

	1	2
1	-0.290964047553239 + 0.000000000000000i	-0.331182602560587 + 0.000000000000000i
2	0.0139605327114863 + 0.229274682642203i	0.00840611116570040 + 0.284400369521363i

3	0.105319522940977 - 0.00130603672270202i	0.170789256021959 - 0.0396389979092825i
4	-0.00504222839112162 - 0.0146563485638448i	-0.0371037316723838 - 0.0637628810754610i
5	0.0369051830080346 + 0.0569650386229188i	-0.0240278872939455 - 0.0260803196869886i

Sebuah bilangan kompleks terdiri dari bilangan riil dan bilangan imajiner. Untuk menghitung bilangan kompleks pada tabel 4.15., maka bilangan tersebut harus diabsolutkan terlebih dahulu dengan menggunakan rumus $absolut = \sqrt{real^2 + imajiner^2}$, sehingga dari hasil tersebut didapatkan bilangan absolut seperti pada tabel 4.14.

Tabel 4.14. Nilai STFT yang sudah diabsolutkan

index	1	2
1	0.2910	0.3312
2	0.2297	0.2845
3	0.1053	0.1753
4	0.0155	0.0738
5	0.0679	0.0355

Pada aplikasi *Matlab*, menentukan letak elemen tertinggi pada suatu matriks adalah dengan sintaks sebagai berikut. Contoh dalam hal ini adalah menggunakan salah satu data rekaman suara laki-laki saat pengucapan huruf “a”.

```
stft(:);
[stft_value,index_no] = max(stft(:));
[I_row, I_col] = ind2sub(size(stft),index_no);
STFT_Value = abs(max(max(stft)));
```

stft adalah matriks STFT yang berukuran (2049 x n), dimana pada penelitian ini, dalam *Matlab*, semua elemen dapat diakses dengan membuat sintaks `stft(:);`.

Untuk mencari letak nilai stft tertinggi berada di elemen keberapa, maka digunakan sintaks `[stft_value, index_no] = max(stft(:));`

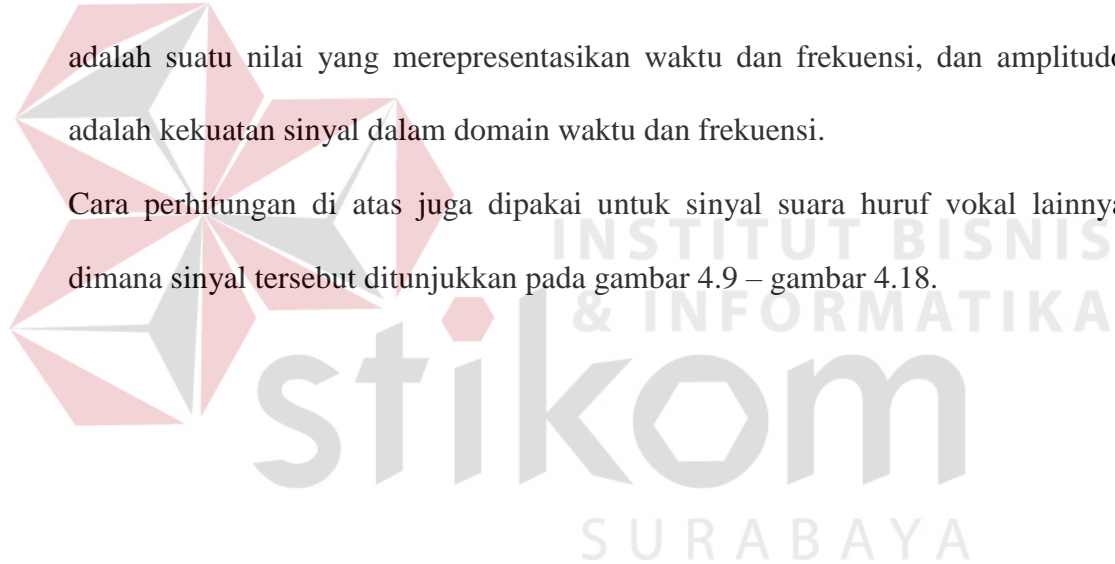
```
[I_row, I_col] = ind2sub(size(stft), index_no);
```

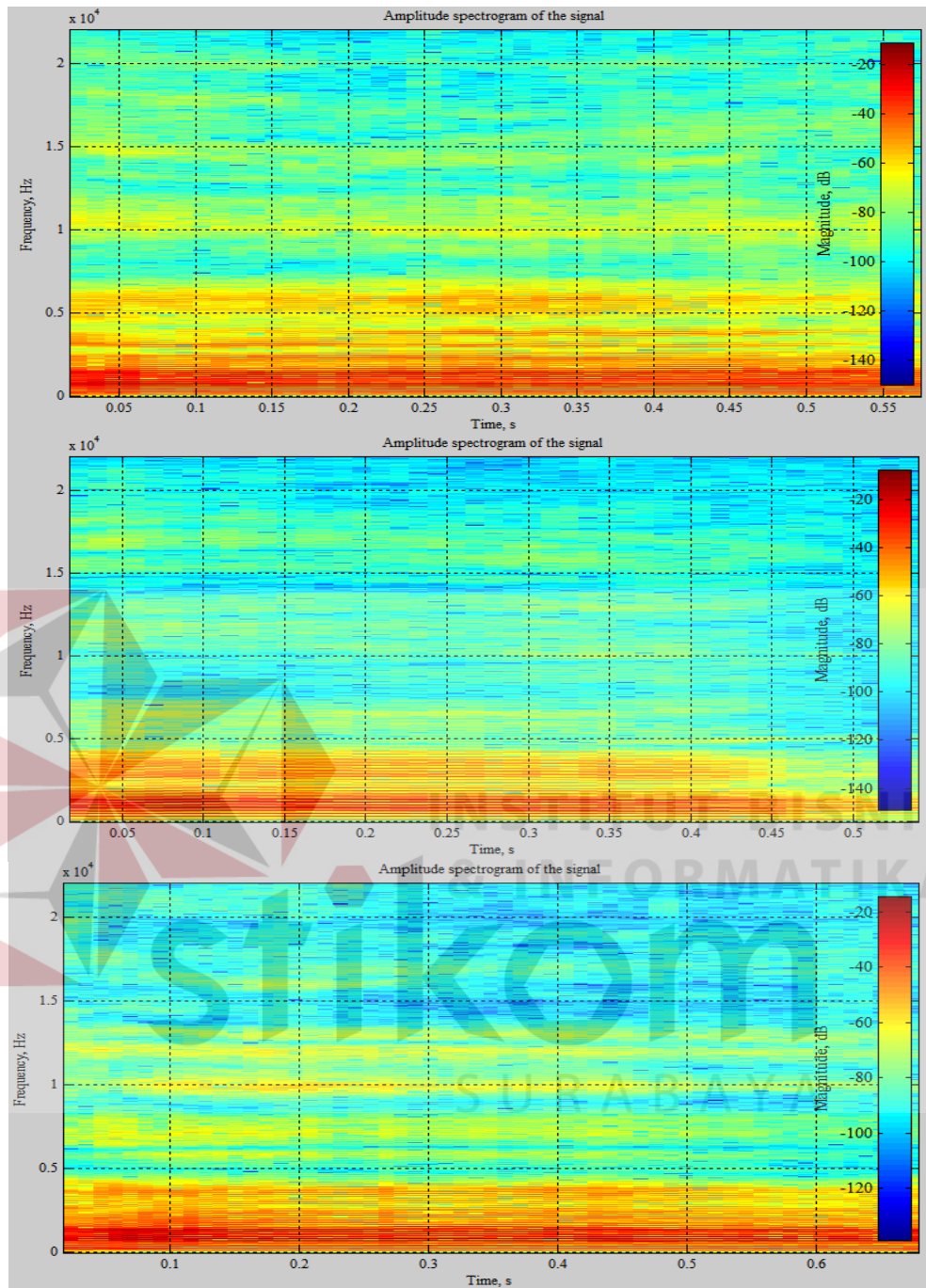
```
STFT_Value = abs(max(max(stft)));
```

I_row dan I_col adalah lokasi nilai STFT tertinggi, dan STFT_Value adalah nilai STFT tertinggi, sehingga nilai yang didapatkan adalah 103.5701. Ukuran matriks

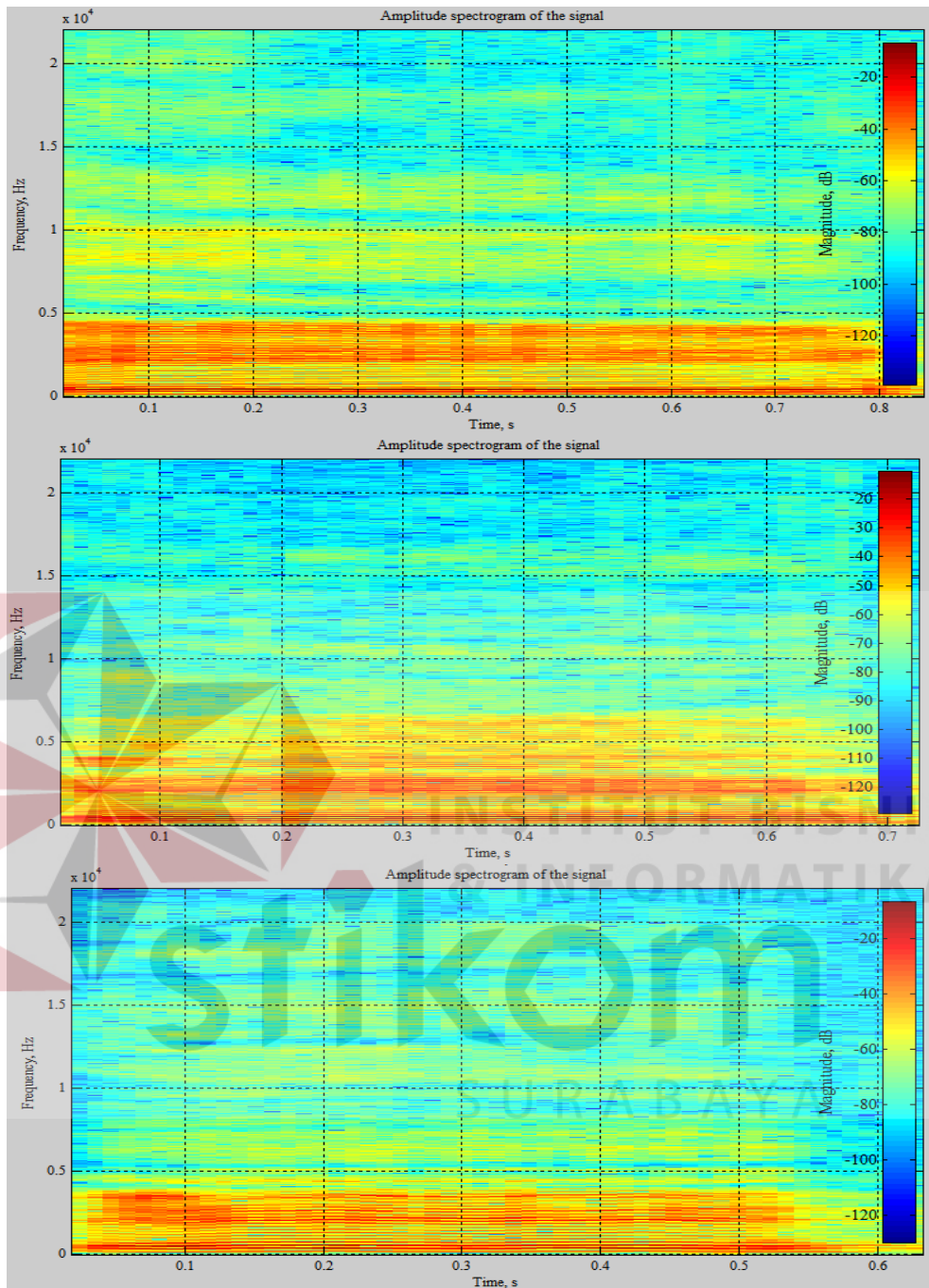
STFT sama dengan ukuran matriks amplitudo yaitu 2049x68 karena nilai STFT adalah suatu nilai yang merepresentasikan waktu dan frekuensi, dan amplitudo adalah kekuatan sinyal dalam domain waktu dan frekuensi.

Cara perhitungan di atas juga dipakai untuk sinyal suara huruf vokal lainnya dimana sinyal tersebut ditunjukkan pada gambar 4.9 – gambar 4.18.

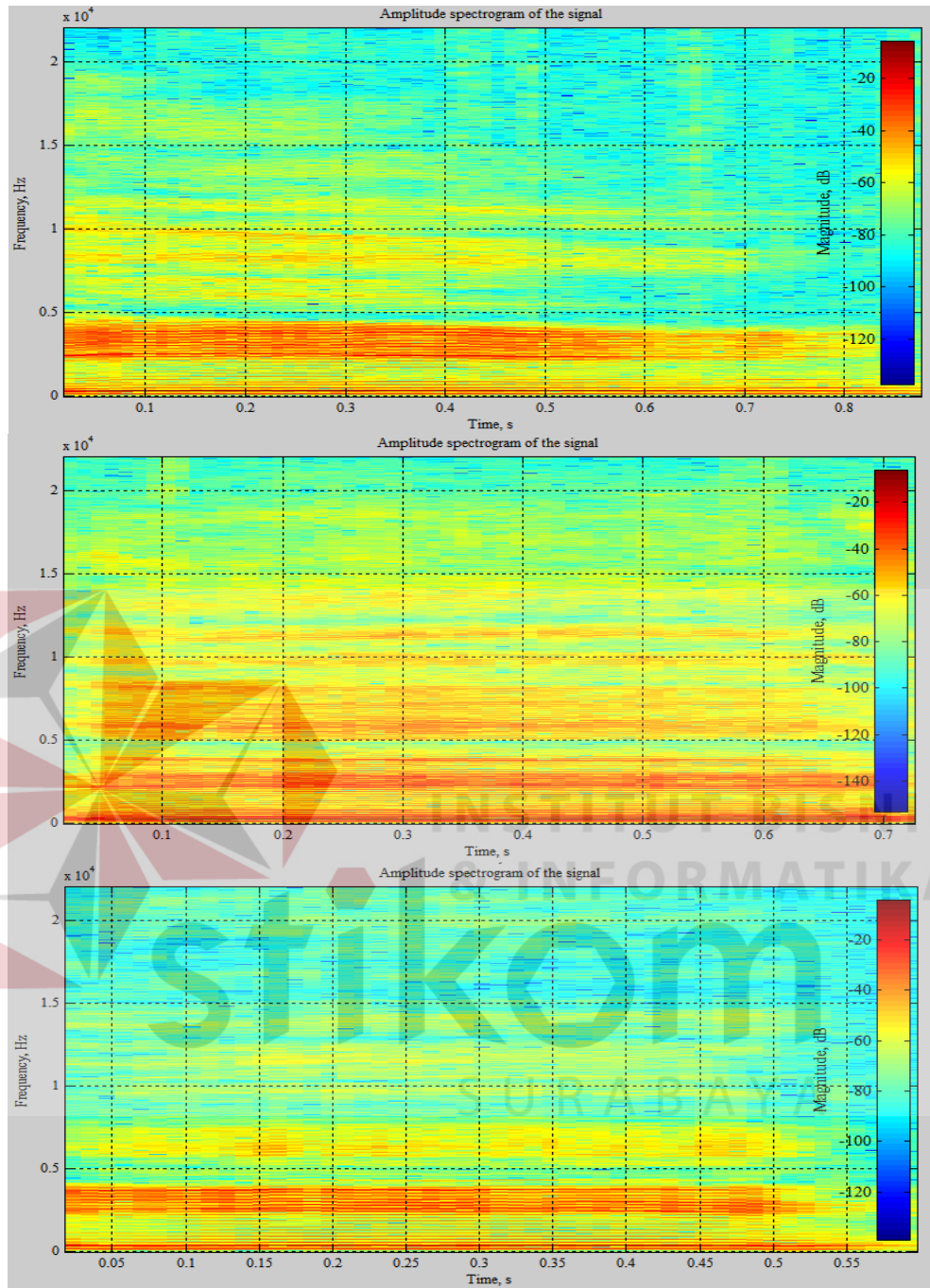




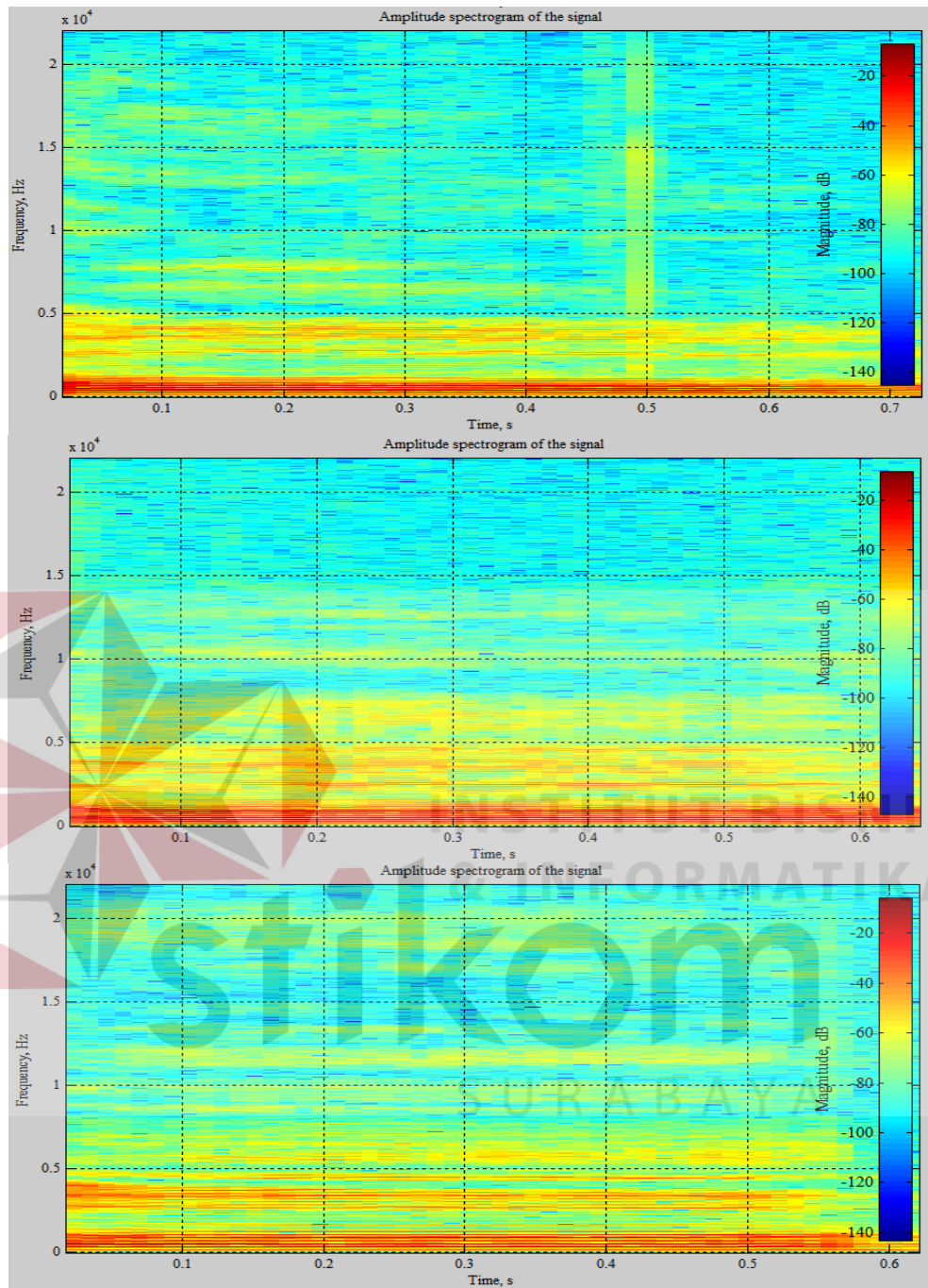
Gambar 4.9. Spektrogram sinyal suara ketiga responden laki-laki saat pengucapan huruf "a"



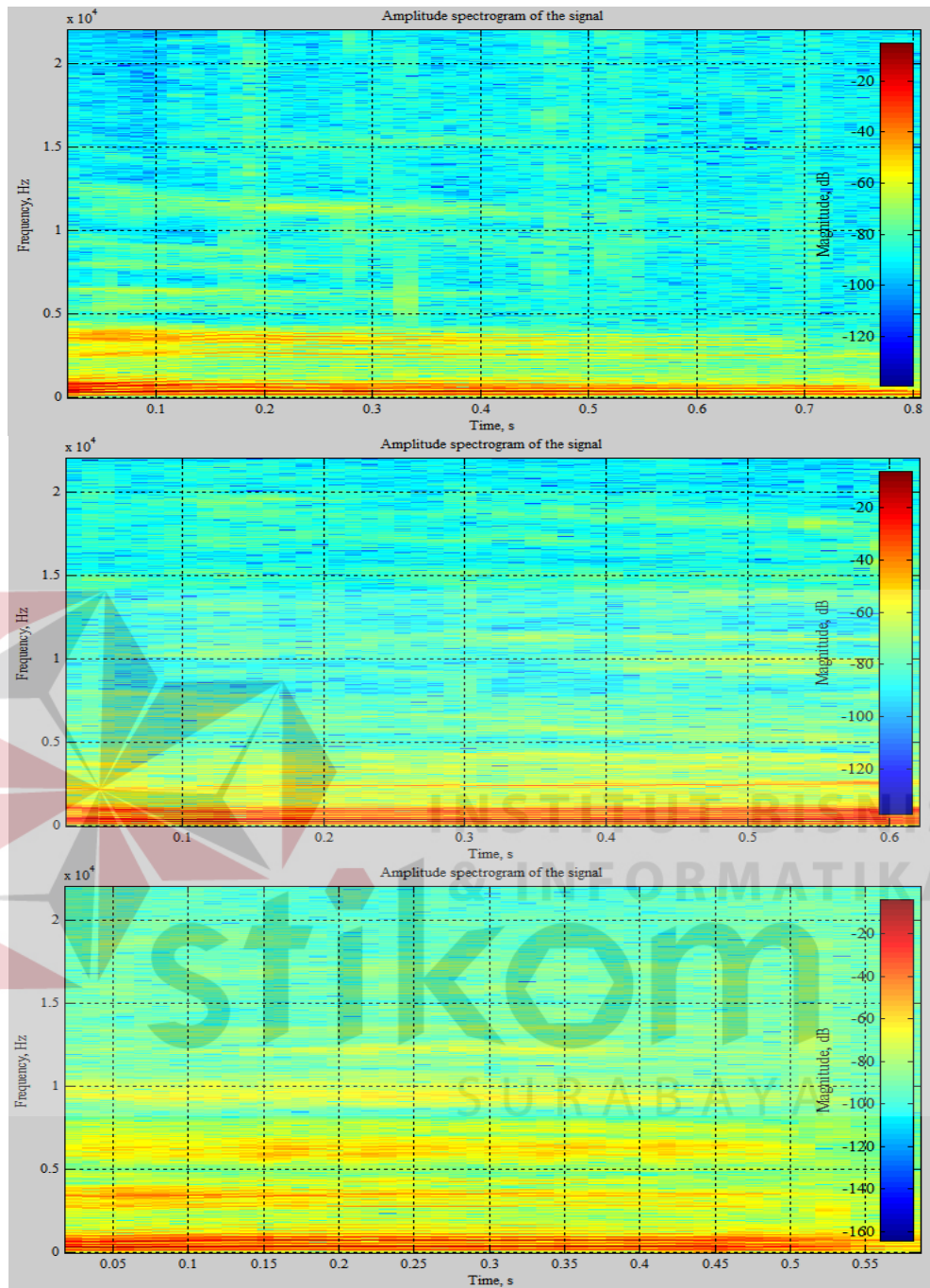
Gambar 4.10. Spektrogram sinyal suara ketiga responden laki-laki saat pengucapan huruf “e”



Gambar 4.11. Spektrogram sinyal suara ketiga responden laki-laki saat pengucapan huruf "i"



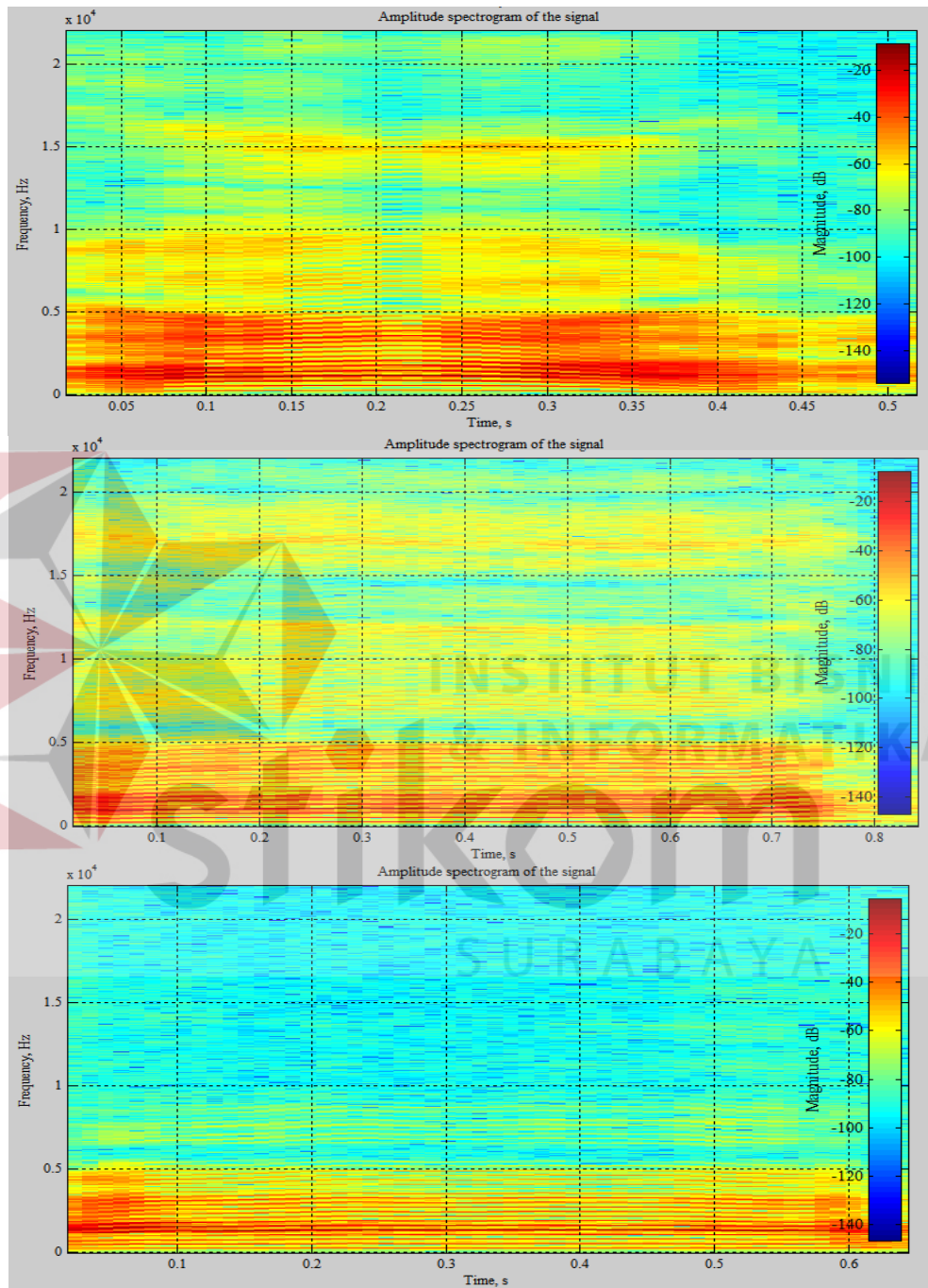
Gambar 4.12. Spektrogram sinyal suara ketiga responden laki-laki saat pengucapan huruf “o”



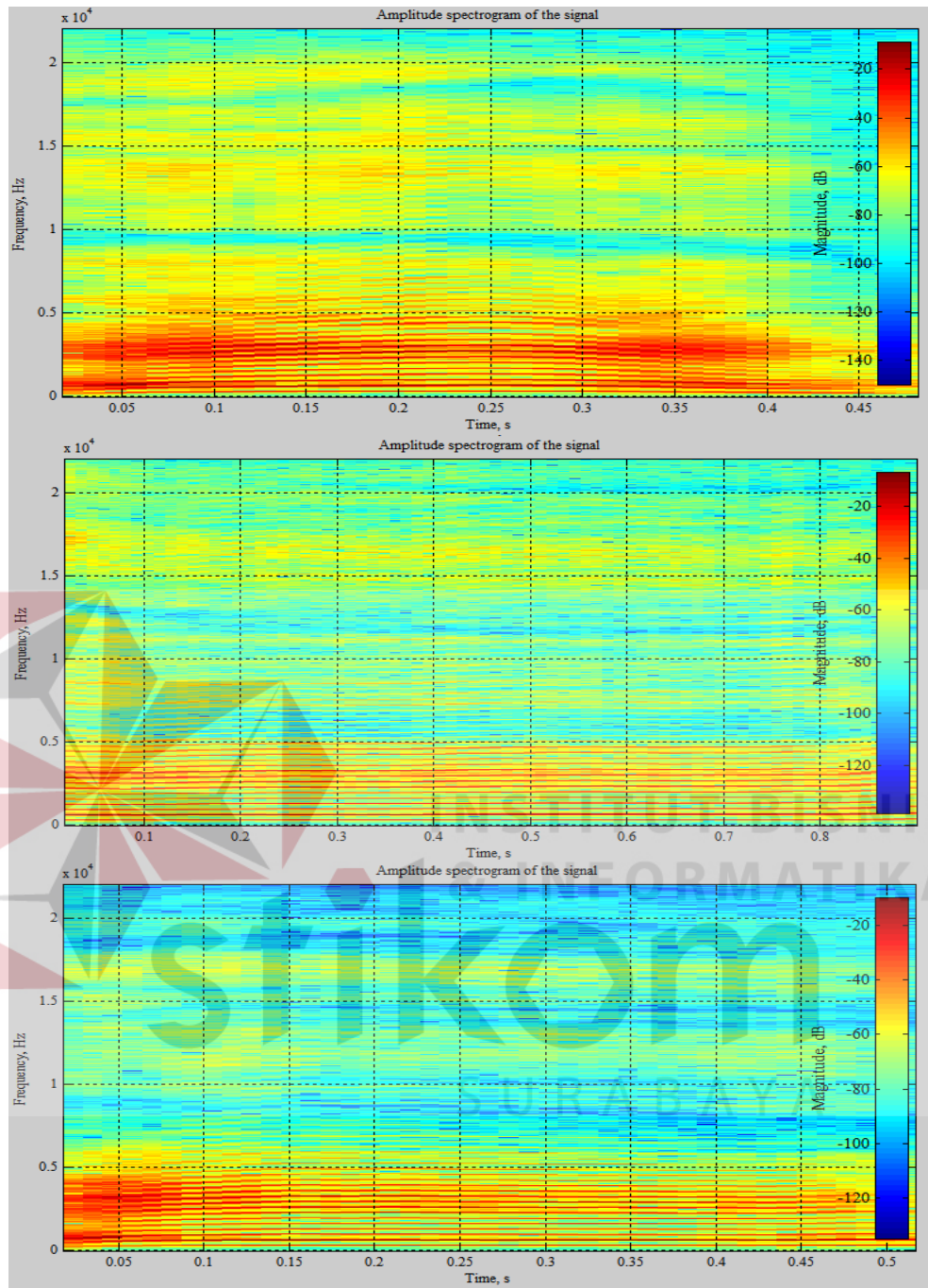
Gambar 4.13. Spektrogram sinyal suara ketiga responden laki-laki saat pengucapan huruf “u”

Gambar 4.9. – gambar 4.13. menunjukkan gambar spektrogram yang berbeda-beda yang dihasilkan oleh 3 orang laki-laki saat pengucapan huruf vokal a, e, i, o, u.

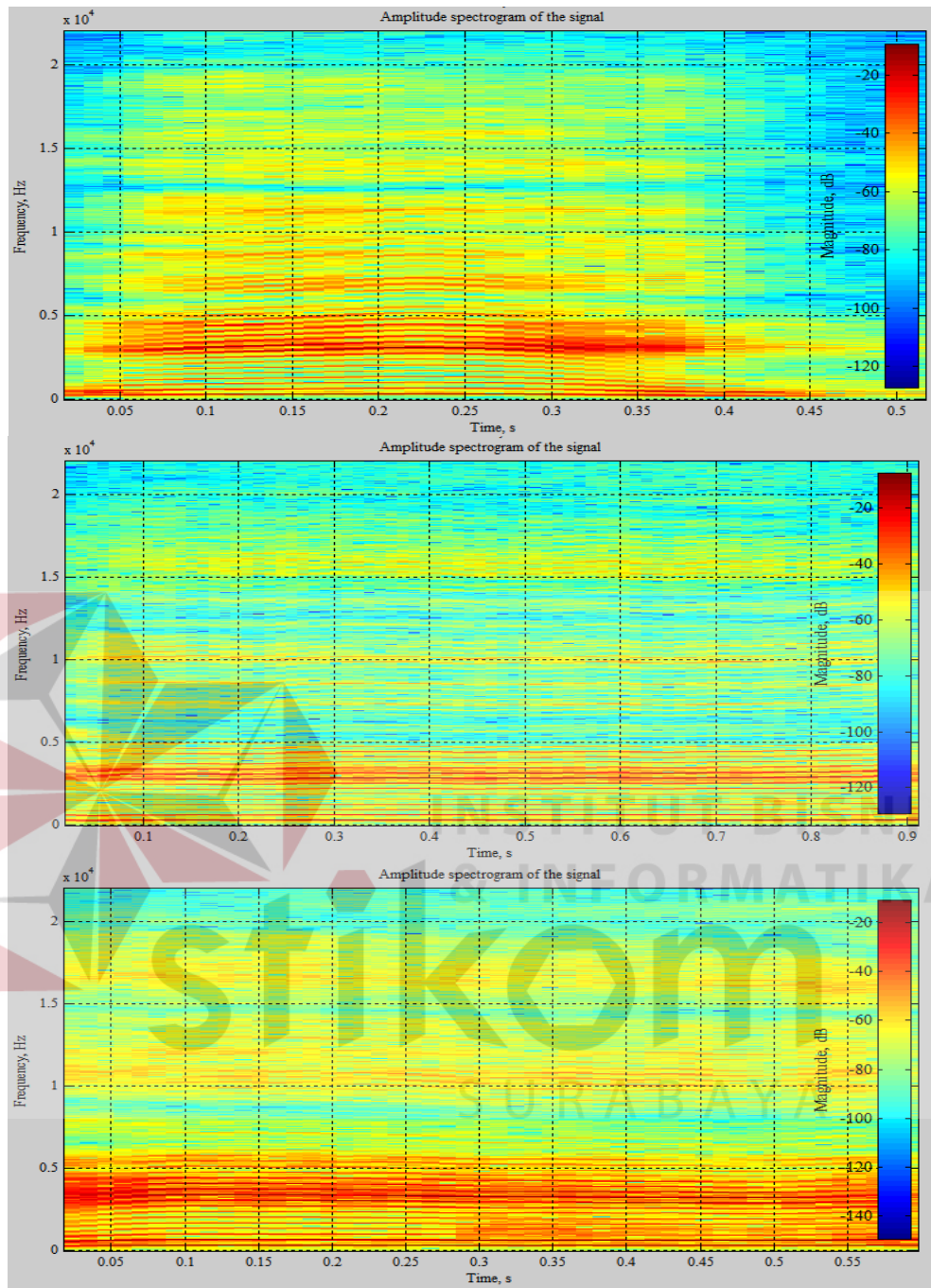
Gambar 4.14 – gambar 4.18. adalah spektrogram sinyal suara yang diambil dari tiga orang responden perempuan pada saat mengucapkan huruf a, e, i, o, u.



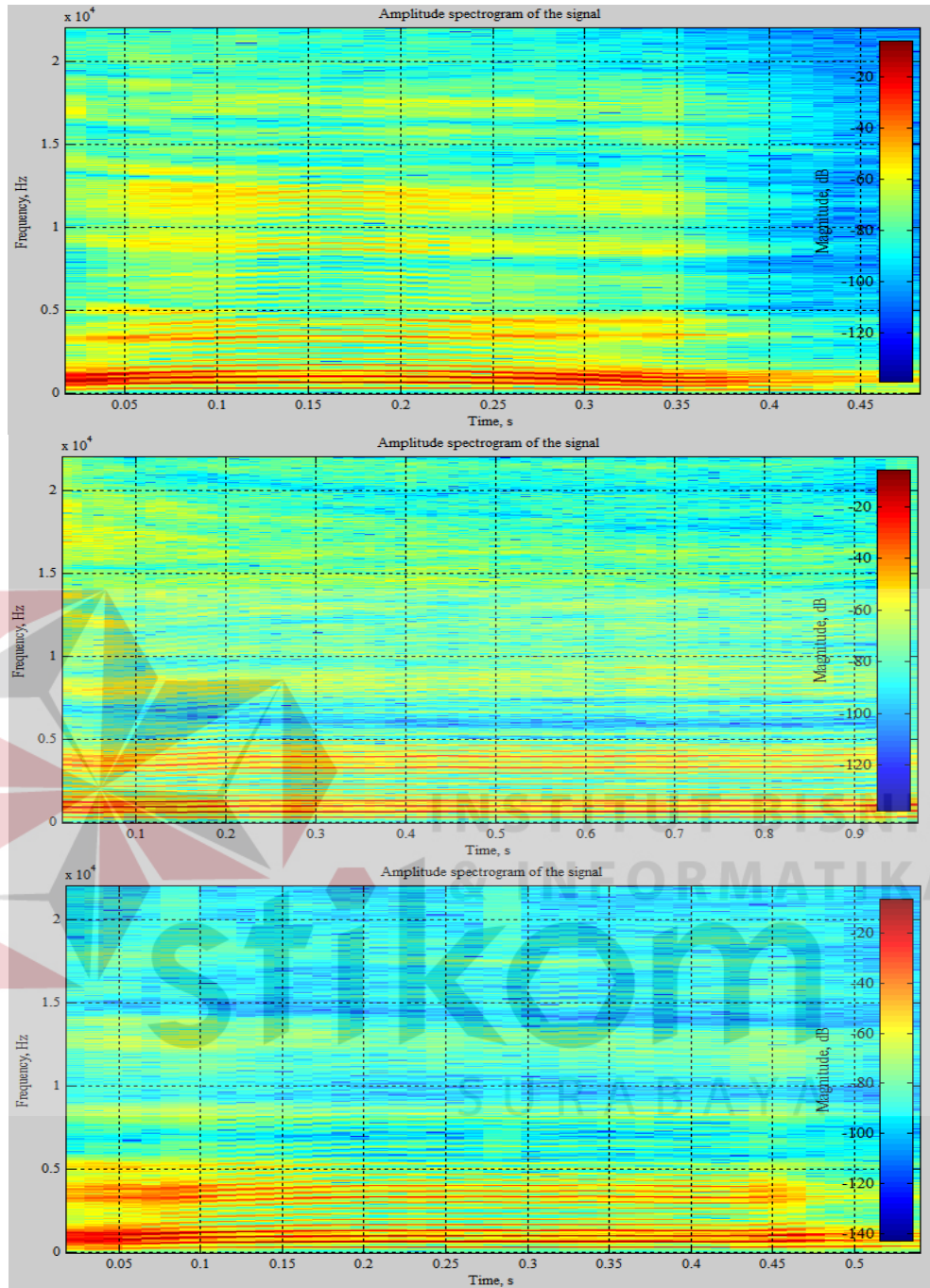
Gambar 4.14. Spektrogram sinyal suara ketiga responden perempuan saat pengucapan huruf “a”



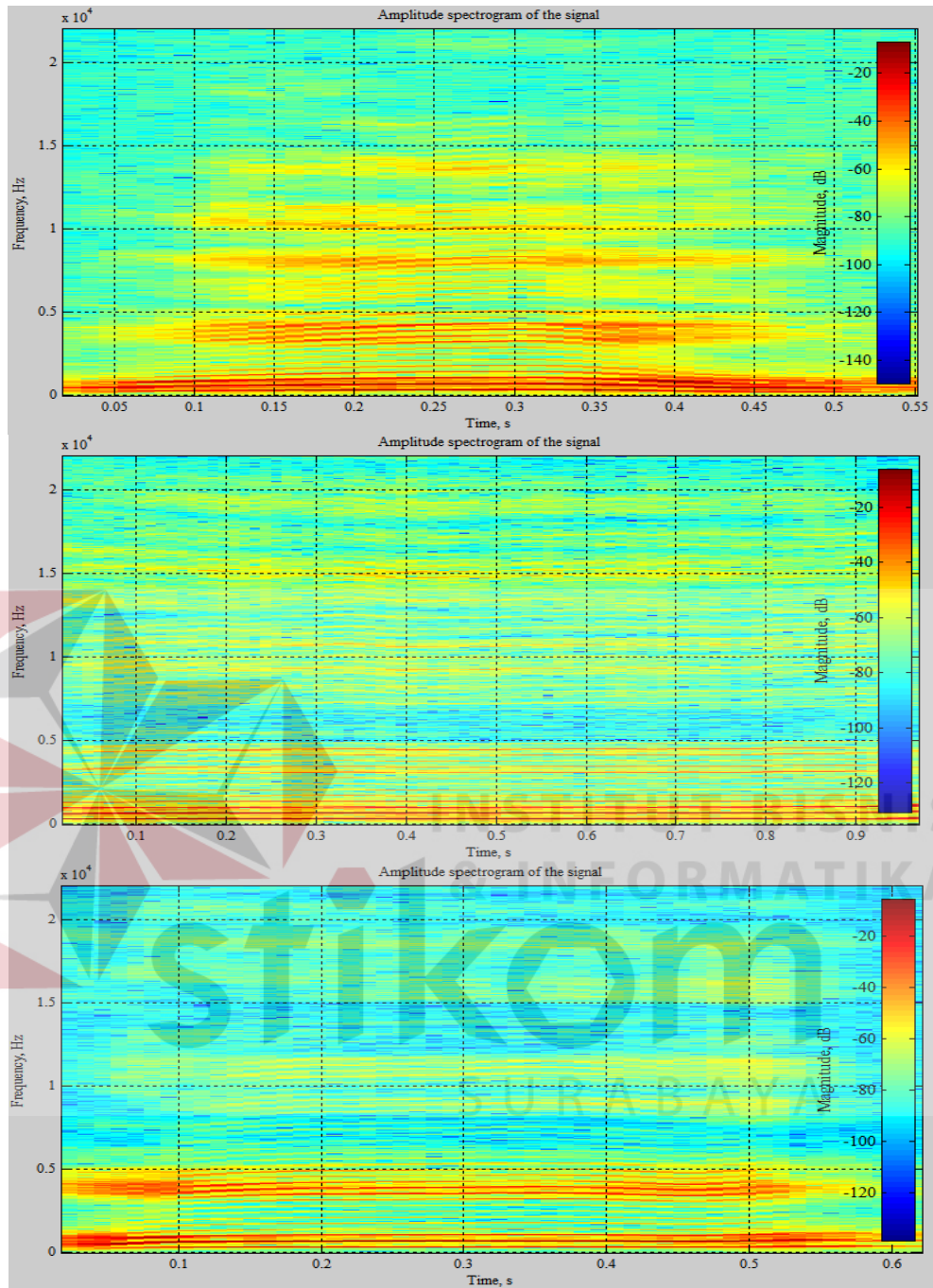
Gambar 4.15. Spektrogram sinyal suara ketiga responden perempuan saat pengucapan huruf “e”



Gambar 4.16. Spektrogram sinyal suara ketiga responden perempuan saat pengucapan huruf "i"



Gambar 4.17. Spektrogram sinyal suara ketiga responden perempuan saat pengucapan huruf “o”



Gambar 4.18. Spektrogram sinyal suara ketiga responden perempuan saat pengucapan huruf “u”

Gambar 4.9. – gambar 4.18. menunjukkan adanya perbedaan yang terjadi antara suara laki-laki dan perempuan, dimana suara laki-laki memiliki jangkauan

frekuensi yang lebih rendah dari suara wanita. Hal ini dikarenakan Karena pada umumnya suara perempuan lebih tinggi 1-2 oktaf dari suara laki-laki.

Nilai STFT_value, amplitudo tertinggi, waktu saat amplitudo tertinggi, dan frekuensi saat amplitudo tertinggi kemudian dikelompokkan ke dalam tabel.

Tabel 4.15. Hasil ekstraksi fitur sinyal suara “a” pada laki-laki

<i>Nama File</i>	<i>Nilai STFT tertinggi</i>	<i>Amplitudo tertinggi (dB)</i>	<i>Saat Amplitudo tertinggi</i>	
			Waktu (s)	Frekuensi (Hz)
<i>a1</i>	193.0197	-8.4734	0.2554	678.2959
<i>a2</i>	211.7397	-7.6694	0.1161	882.8613
<i>a3</i>	206.6431	-7.8810	0.0580	925.9277
<i>a4</i>	146.0479	-10.8955	0.1858	893.6279
<i>a5</i>	152.8552	-10.4998	0.2902	678.2959
<i>a6</i>	195.3452	-8.3693	0.1858	732.1289
<i>a7</i>	98.4328	-14.3226	0.1393	829.0283
<i>a8</i>	142.9596	-11.0811	0.1161	785.9619
<i>a9</i>	101.9875	-14.0145	0.1625	818.2617
<i>a10</i>	274.9236	-5.4012	0.0929	807.4951
<i>a11</i>	103.5701	-13.8807	0.1625	1313.5254
<i>a12</i>	203.2935	-8.0229	0.0232	742.8955
<i>a13</i>	159.0506	-10.1547	0.3483	872.0947
<i>a14</i>	146.9037	-10.8447	0.0813	915.1611
<i>a15</i>	162.9476	-9.9444	0.3019	764.4287

Tabel 4.16. Hasil ekstraksi fitur sinyal suara “a” pada perempuan

<i>Nama File</i>	<i>Nilai STFT tertinggi</i>	<i>Amplitudo tertinggi (dB)</i>	<i>Saat Amplitudo tertinggi</i>	
			Waktu (s)	Frekuensi (Hz)
<i>a1</i>	129.7541	-11.9230	0.0580	979.7607
<i>a2</i>	212.3353	-7.6450	0.1045	1001.2939

<i>a3</i>	247.8205	-6.3027	0.0697	968.9941
<i>a4</i>	117.5819	-12.7786	0.0929	204.5654
<i>a5</i>	116.5292	-12.8567	0.3520	1015.6250
<i>a6</i>	152.8072	-10.5025	0.6720	773.4375
<i>a7</i>	117.5024	-12.7845	0.3019	882.8613
<i>a8</i>	79.9360	-16.1306	0.0580	1475.0244
<i>a9</i>	107.9746	-13.5190	0.6618	775.1953
<i>a10</i>	140.9873	-11.2018	0.6502	850.5615
<i>a11</i>	185.7544	-8.8066	0.0697	1152.0264
<i>a12</i>	225.4146	-7.1258	0.0580	1216.6260
<i>a13</i>	108.2563	-13.4963	0.1161	1108.9600
<i>a14</i>	148.8902	-10.7281	0.0929	1012.0605
<i>a15</i>	156.6096	-10.2890	0.1161	1087.4268

Tabel 4.15. dan tabel 4.16. menunjukkan hasil ekstraksi ciri oleh metode STFT. Terdapat perbedaan pada suara laki-laki dan suara perempuan dimana pada suara perempuan nilai frekuensi lebih tinggi dari suara laki-laki. Jika merujuk pada tabel 2.1. di BAB II, maka dapat dilihat bahwa nilai frekuensi yang didapat dari hasil STFT mendekati dari frekuensi *range* vokal pada tabel 2.1. Terdapat perbedaan antara nilai frekuensi STFT lebih besar dari frekuensi *range* vokal tersebut. Hal ini disebabkan karena volume suara yang besar saat perekaman suara sehingga menyebabkan frekuensi yang besar juga.

Adapun nilai-nilai yang diekstraksi pada penelitian ini adalah nilai STFT, nilai rata-rata amplitudo pada sebuah sinyal, nilai frekuensi dan nilai waktu saat amplitudo tertinggi seperti terlihat pada tabel. Nilai STFT merupakan bilangan kompleks perlu di-absolutkan, sementara nilai amplitudo adalah nilai dalam satuan desibel.

Nilai-nilai pada tabel 4.15. dan 4.16. masih merupakan nilai yang tidak seragam. Contohnya adalah jarak antara nilai waktu yang kecil dan nilai frekuensi yang besar. Perbedaan ini membuat nilai tersebut tidak bisa dijadikan *input* pada *backpropagation*. Sehingga nilai-nilai tersebut harus diseragamkan terlebih dahulu atau dinormalisasikan menjadi sebuah bilangan antara 0 dan 1. Adapun contoh membuat rumus normalisasi pada Matlab adalah

```
normSA = dBSA - min(dBSA(:));
normSA = normSA ./ max(normSA(:));
```

Nilai normSA adalah suatu matriks yang berisi hasil normalisasi. Dan nilai dBSA adalah suatu matriks yang belum ternormalisasi. Nilai tersebut akan dikurangi dengan nilai yang lebih kecil antara elemen ke i baru, dengan elemen sebelumnya. Sehingga nilai normSA akan terus diperbaharui selama belum mencapai elemen terakhir pada matriks. Nilai setiap elemen pada matriks normSA yang telah diperbaharui kemudian akan dibagi dengan sebuah nilai dari elemen terbesar yang ada matriks tersebut. Sehingga didapatkanlah sebuah matriks baru hasil ternormalisasi.

Contohnya adalah menormalisasi nilai frekuensi. Nilai frekuensi yang dinormalisasi adalah nilai frekuensi pada satu data rekaman sinyal suara. Adapun nilai terbesar dari hasil normalisasi selalu menunjukkan angka 1, dan nilai terkecil adalah nilai 0, hal ini dikarenakan data dinormalisasi menjadi nilai antara 0 sampai 1. Dalam hal ini, penggunaan nilai maksimal atau nilai terbesar tidak berlaku untuk nilai STFT dan nilai amplitudo, karena pada perhitungan sebelumnya nilai yang digunakan adalah nilai terbesar dari matriks STFT, dan nilai terbesar dari nilai amplitudo. Sehingga untuk alternatifnya adalah merata-rata

nilai yang sudah dinormalisasi untuk mendapatkan nilai yang berbeda setiap rekaman sinyal suara.

Contohnya nilai STFT pada salah satu sinyal suara rekaman responden laki-laki saat pengucapan huruf “a” seperti pada tabel 4.16. Nilai dari matriks STFT untuk satu data rekaman suara pengucapan yang sudah ternormalisasi tersebut akan dirata-rata. Sehingga nilai rata-rata STFT untuk satu data rekaman suara salah satu responden laki-laki saat pengucapan huruf “a” adalah 0.0061

Untuk nilai amplitudo juga mengalami proses normalisasi karena nilai dari matriks amplitudo masih bernilai negatif seperti pada tabel 4.15 dan tabel 4.16. Namun saat dilakukan normalisasi, nilai amplitudo terbesar selalu bernilai 1, sehingga nilai hasil normalisasi amplitudo akan dirata-rata. Adapun nilai yang dirata-rata adalah nilai pada matriks amplitudo pada tiap data rekaman sinyal suara.

Setelah proses normalisasi, maka nilai-nilai yang ada pada matriks tersebut akan dirata-rata. Nilai rata-rata yang didapatkan pada matriks amplitudo yang sudah ternormalisasi dari sinyal suara salah satu responden laki-laki saat pengucapan huruf “a” adalah 0.4819.

Agar semua nilai seragam, maka nilai waktu pun harus dinormalisasi juga. Adapun nilai waktu dan frekuensi yang digunakan adalah tetap pada nilai waktu dan frekuensi saat amplitudo tertinggi, namun nilai waktu dan frekuensi tersebut adalah nilai yang sudah ternormalisasi.

Seluruh nilai-nilai STFT, nilai rata-rata amplitudo, waktu, dan frekuensi yang sudah ternormalisasi dikumpulkan menjadi satu tabel untuk digolongkan berdasarkan jenis kelamin dan huruf vokalnya. Seperti tertera pada tabel 4.17.

sampai tabel 4.21. untuk suara laki-laki, dan tabel 4.22. sampai tabel 4.26. untuk suara perempuan

Tabel 4.17. Hasil ekstraksi ciri sinyal suara “a” pada laki-laki yang sudah dinormalisasi

<i>Nama File</i>	<i>Nilai STFT</i>	<i>Rata-rata amplitudo</i>	<i>Saat Amplitudo Maksimal</i>	
			waktu	frekuensi
<i>a1</i>	0.0061	0.4819	0.2985	0.0308
<i>a2</i>	0.0051	0.4604	0.1818	0.0400
<i>a3</i>	0.0050	0.4879	0.0545	0.0420
<i>a4</i>	0.0067	0.4459	0.2373	0.0405
<i>a5</i>	0.0084	0.4713	0.6571	0.0308
<i>a6</i>	0.0068	0.4907	0.2333	0.0332
<i>a7</i>	0.0125	0.4805	0.1786	0.0376
<i>a8</i>	0.0086	0.5221	0.1702	0.0356
<i>a9</i>	0.0125	0.5276	0.2667	0.0371
<i>a10</i>	0.0054	0.4717	0.1034	0.0366
<i>a11</i>	0.0137	0.5296	0.1846	0.0596
<i>a12</i>	0.0071	0.4707	0.0000	0.0337
<i>a13</i>	0.0077	0.5120	0.3590	0.0396
<i>a14</i>	0.0106	0.5207	0.1613	0.0415
<i>a15</i>	0.0077	0.5194	0.3038	0.0347

Tabel 4.18. Hasil ekstraksi ciri sinyal suara “e” pada laki-laki yang sudah dinormalisasi

<i>Nama File</i>	<i>Nilai STFT</i>	<i>Rata-rata amplitudo</i>	<i>Saat Amplitudo Maksimal</i>	
			waktu	frekuensi
<i>e1</i>	0.0074	0.5181	0.6000	0.0205
<i>e2</i>	0.0056	0.4693	0.6731	0.0176
<i>e3</i>	0.0074	0.5778	0.0612	0.0244
<i>e4</i>	0.0056	0.4717	0.3810	0.0220
<i>e5</i>	0.0042	0.4685	0.5400	0.0181
<i>e6</i>	0.0048	0.4544	0.4559	0.0195

<i>e7</i>	0.0057	0.4974	0.2571	0.0210
<i>e8</i>	0.0092	0.4768	0.4000	0.0234
<i>e9</i>	0.0063	0.5023	0.6923	0.0200
<i>e10</i>	0.0062	0.5261	0.2982	0.0225
<i>e11</i>	0.0112	0.5355	0.6056	0.0195
<i>e12</i>	0.0045	0.5047	0.2222	0.0195
<i>e13</i>	0.0043	0.4759	0.2143	0.0215
<i>e14</i>	0.0141	0.5220	0.0588	0.0264
<i>e15</i>	0.0093	0.5087	0.4444	0.0195

Tabel 4.19. Hasil ekstraksi ciri sinyal suara “i” pada laki-laki yang sudah dinormalisasi

<i>Nama File</i>	<i>Nilai STFT</i>	<i>Rata-rata amplitudo</i>	<i>Saat Amplitudo Maksimal</i>	
			waktu	frekuensi
<i>i1</i>	0.0068	0.5519	0.6324	0.0127
<i>i2</i>	0.0041	0.4991	0.4694	0.0171
<i>i3</i>	0.0046	0.5121	0.1154	0.0093
<i>i4</i>	0.0049	0.5251	0.1091	0.0156
<i>i5</i>	0.0055	0.4694	0.1667	0.0112
<i>i6</i>	0.0060	0.4897	0.1750	0.0186
<i>i7</i>	0.0045	0.4989	0.3014	0.0156
<i>i8</i>	0.0057	0.5764	0.1833	0.0127
<i>i9</i>	0.0053	0.4751	0.3143	0.0146
<i>i10</i>	0.0037	0.5169	0.1186	0.0151
<i>i11</i>	0.0079	0.5551	0.0000	0.0205
<i>i12</i>	0.0106	0.5109	0.0469	0.0098
<i>i13</i>	0.0045	0.4947	0.9571	0.0146
<i>i14</i>	0.0136	0.5441	0.2759	0.1240
<i>i15</i>	0.0073	0.4904	0.3269	0.0127

Tabel 4.20. Hasil ekstraksi ciri sinyal suara “o” pada laki-laki yang sudah dinormalisasi

<i>Nama File</i>	<i>Nilai STFT</i>	<i>Rata-rata amplitudo</i>	<i>Saat Amplitudo Maksimal</i>	
			waktu	frekuensi
<i>o1</i>	0.0042	0.4419	0.5541	0.0210
<i>o2</i>	0.0046	0.4818	0.0784	0.0234
<i>o3</i>	0.0038	0.4864	0.0545	0.0200
<i>o4</i>	0.0046	0.4297	0.1633	0.0205
<i>o5</i>	0.0068	0.4465	0.1818	0.0181
<i>o6</i>	0.0044	0.4190	0.5432	0.0186
<i>o7</i>	0.0045	0.4469	0.2500	0.0195
<i>o8</i>	0.0063	0.4772	0.2264	0.0415
<i>o9</i>	0.0044	0.5052	0.4667	0.0210
<i>o10</i>	0.0044	0.4408	0.2083	0.0234
<i>o11</i>	0.0070	0.4467	0.0625	0.0244
<i>o12</i>	0.0039	0.4582	0.2278	0.0195
<i>o13</i>	0.0054	0.4649	0.3797	0.0220
<i>o14</i>	0.0054	0.4442	0.1714	0.0396
<i>o15</i>	0.0056	0.4447	0.3125	0.0190

Tabel 4.21. Hasil ekstraksi ciri sinyal suara “u” pada laki-laki yang sudah dinormalisasi

<i>Nama File</i>	<i>Nilai STFT</i>	<i>Rata-rata amplitudo</i>	<i>Saat Amplitudo Maksimal</i>	
			waktu	frekuensi
<i>u1</i>	0.0035	0.4521	0.6290	0.0146
<i>u2</i>	0.0028	0.5142	0.2292	0.0171
<i>u3</i>	0.0028	0.4490	0.1091	0.0117
<i>u4</i>	0.0045	0.4339	0.8246	0.0151
<i>u5</i>	0.0057	0.4625	0.5690	0.0161
<i>u6</i>	0.0037	0.4207	0.3816	0.0195
<i>u7</i>	0.0027	0.4242	0.0896	0.0171
<i>u8</i>	0.0049	0.4517	0.6863	0.0186
<i>u9</i>	0.0048	0.4332	0.4524	0.0156

<i>u10</i>	0.0042	0.5202	0.8302	0.0156
<i>u11</i>	0.0046	0.4999	0.3165	0.0142
<i>u12</i>	0.0043	0.4585	0.0500	0.0205
<i>u13</i>	0.0038	0.4814	0.1452	0.0151
<i>u14</i>	0.0035	0.4282	0.0968	0.0205
<i>u15</i>	0.0050	0.4622	0.6769	0.0127

Berdasarkan tabel 4.17. – tabel 4.21. nilai-nilai yang ada sudah menjadi seragam, yaitu berkisar antara 0 sampai 1.

Tabel 4.22. sampai tabel 4.26. adalah hasil ekstraksi ciri sinyal suara pada perempuan:

Tabel 4.22. Hasil ekstraksi ciri sinyal suara “a” pada perempuan yang sudah dinormalisasi

<i>Nama File</i>	<i>Nilai STFT</i>	<i>Rata-rata amplitudo</i>	<i>Saat Amplitudo Maksimal</i>	
			waktu	frekuensi
<i>a1</i>	0.0068	0.5473	0.2857	0.0522
<i>a2</i>	0.0064	0.5506	0.1000	0.0454
<i>a3</i>	0.0046	0.4567	0.2264	0.0601
<i>a4</i>	0.0099	0.5817	0.4074	0.0591
<i>a5</i>	0.0146	0.5784	0.6923	0.1270
<i>a6</i>	0.0129	0.5367	0.7308	0.0967
<i>a7</i>	0.0113	0.4501	0.6000	0.0400
<i>a8</i>	0.0092	0.4921	0.0536	0.0669
<i>a9</i>	0.0111	0.4634	0.6875	0.0352
<i>a10</i>	0.0101	0.5044	0.7606	0.0386
<i>a11</i>	0.0063	0.4258	0.1600	0.0522
<i>a12</i>	0.0064	0.5284	0.7436	0.0576
<i>a13</i>	0.0097	0.6233	0.1509	0.0503
<i>a14</i>	0.0102	0.5412	0.2500	0.0459
<i>a15</i>	0.0102	0.4992	0.2667	0.0493

Tabel 4.23. Hasil ekstraksi ciri sinyal suara “e” pada perempuan yang sudah dinormalisasi

<i>Nama File</i>	<i>Nilai STFT</i>	<i>Rata-rata amplitudo</i>	<i>Saat Amplitudo Maksimal</i>	
			waktu	frekuensi
<i>e1</i>	0.0075	0.5562	0.4872	0.0313
<i>e2</i>	0.0056	0.5112	0.0400	0.0288
<i>e3</i>	0.0083	0.4709	0.2619	0.0298
<i>e4</i>	0.0065	0.5474	0.3951	0.0327
<i>e5</i>	0.0142	0.5139	0.3000	0.0571
<i>e6</i>	0.0203	0.5803	0.5200	0.0596
<i>e7</i>	0.0085	0.4637	0.5323	0.0190
<i>e8</i>	0.0090	0.5084	0.0323	0.0254
<i>e9</i>	0.0072	0.4943	0.2692	0.0181
<i>e10</i>	0.0087	0.4969	0.3286	0.0176
<i>e11</i>	0.0057	0.4774	0.3226	0.0098
<i>e12</i>	0.0090	0.4711	0.2286	0.0264
<i>e13</i>	0.0098	0.5805	0.4211	0.0283
<i>e14</i>	0.0135	0.4727	0.1667	0.0190
<i>e15</i>	0.0079	0.5184	0.4146	0.0327

Tabel 4.24. Hasil ekstraksi ciri sinyal suara “i” pada perempuan yang sudah dinormalisasi

<i>Nama File</i>	<i>Nilai STFT</i>	<i>Rata-rata amplitudo</i>	<i>Saat Amplitudo Maksimal</i>	
			waktu	frekuensi
<i>i1</i>	0.0084	0.5143	0.4286	0.1416
<i>i2</i>	0.0072	0.4995	0.9737	0.0151
<i>i3</i>	0.0112	0.5813	0.7143	0.1475
<i>i4</i>	0.0155	0.5659	0.3143	0.1265
<i>i5</i>	0.0128	0.5486	0.9286	0.0601
<i>i6</i>	0.0268	0.6026	0.5385	0.5474
<i>i7</i>	0.0072	0.4886	0.8222	0.0176
<i>i8</i>	0.0123	0.4706	0.6111	0.0190
<i>i9</i>	0.0060	0.4488	0.7324	0.0181

<i>i10</i>	0.0079	0.5345	0.8289	0.0098
<i>i11</i>	0.0038	0.4192	0.5172	0.0098
<i>i12</i>	0.0096	0.4766	0.8182	0.0273
<i>i13</i>	0.0042	0.5287	0.2727	0.0098
<i>i14</i>	0.0045	0.4437	0.2083	0.0112
<i>i15</i>	0.0043	0.4688	0.2564	0.0103

Tabel 4.25. Hasil ekstraksi ciri sinyal suara “o” pada perempuan yang sudah dinormalisasi

<i>Nama File</i>	<i>Nilai STFT</i>	<i>Rata-rata amplitudo</i>	<i>Saat Amplitudo Maksimal</i>	
			waktu	frekuensi
<i>o1</i>	0.0043	0.4676	0.3077	0.0317
<i>o2</i>	0.0055	0.4783	0.6790	0.0303
<i>o3</i>	0.0049	0.4403	0.6136	0.0308
<i>o4</i>	0.0054	0.5248	0.0500	0.0317
<i>o5</i>	0.0078	0.4841	0.1818	0.0552
<i>o6</i>	0.0079	0.5033	0.6522	0.1294
<i>o7</i>	0.0058	0.4256	0.4407	0.0469
<i>o8</i>	0.0038	0.4043	0.7069	0.0376
<i>o9</i>	0.0067	0.4113	0.0000	0.0264
<i>o10</i>	0.0051	0.4327	0.2877	0.0186
<i>o11</i>	0.0054	0.4097	0.2609	0.0415
<i>o12</i>	0.0050	0.4482	0.4211	0.0249
<i>o13</i>	0.0066	0.5277	0.1905	0.0371
<i>o14</i>	0.0055	0.4251	0.0370	0.0430
<i>o15</i>	0.0070	0.4733	0.0294	0.0439

Tabel 4.26. Hasil ekstraksi ciri sinyal suara “u” pada perempuan yang sudah dinormalisasi

<i>Nama File</i>	<i>Nilai STFT</i>	<i>Rata-rata amplitudo</i>	<i>Saat Amplitudo Maksimal</i>	
			waktu	frekuensi
<i>u1</i>	0.0045	0.5172	0.4889	0.0327

u_2	0.0056	0.4887	0.3457	0.0156
u_3	0.0041	0.4574	0.4118	0.0322
u_4	0.0056	0.5518	0.6250	0.0332
u_5	0.0047	0.4647	0.5294	0.0610
u_6	0.0071	0.5325	0.6538	0.1035
u_7	0.0045	0.4242	0.7647	0.0210
u_8	0.0038	0.4272	0.0816	0.0176
u_9	0.0040	0.4209	0.6563	0.0181
u_{10}	0.0052	0.4247	0.5362	0.0186
u_{11}	0.0036	0.3349	0.6522	0.0190
u_{12}	0.0035	0.4233	0.1064	0.0269
u_{13}	0.0046	0.4665	0.0816	0.0093
u_{14}	0.0047	0.4531	0.1053	0.0220
u_{15}	0.0036	0.4249	0.0952	0.0117

Nilai-nilai pada Tabel 4.21. sampai tabel 4.30. nantinya akan dijadikan sebagai nilai masukan untuk data pelatihan pada jaringan saraf tiruan *backpropagation*.

4.4. Pelatihan dan pengujian *backpropagation*

Pada bagian pelatihan di inputkan data hasil ekstraksi ciri STFT dimana input adalah fitur dari huruf a-e-i-o-u (masing-masing huruf 4 fitur suara) sehingga jumlah *neuron input* adalah sebanyak 20 ditambah 1 *neuron bias*. Dan data yang akan dilatihkan adalah sebanyak jumlah responden yaitu 30 data. Seperti terlihat pada tabel 4.31.

Tabel 4.27. *Neuron input pada backpropagation*

	Huruf vokal																			
	a				e				i				o				u			
Neuron Input	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20
Nilai	Nilai STFT	Amplitudo	Waktu	Frekuensi	Nilai STFT	Amplitudo	Waktu	Frekuensi	Nilai STFT	Amplitudo	Waktu	Frekuensi	Nilai STFT	Amplitudo	Waktu	Frekuensi	Nilai STFT	Amplitudo	Waktu	Frekuensi
Data training 1	0.0068	0.5473	0.2857	0.0522	0.0075	0.5562	0.4872	0.0313	0.0084	0.5143	0.4286	0.1416	0.0043	0.4676	0.3077	0.0317	0.0045	0.5172	0.4889	0.0327
Data training 2	0.0064	0.5506	0.1	0.0454	0.0056	0.5112	0.04	0.0288	0.0072	0.4995	0.9737	0.0151	0.0055	0.4783	0.679	0.0303	0.0056	0.4887	0.3457	0.0156
Data training 3	0.0046	0.4567	0.2264	0.0601	0.0083	0.4709	0.2619	0.0298	0.0112	0.5813	0.7143	0.1475	0.0049	0.4403	0.6136	0.0308	0.0041	0.4574	0.4118	0.0322
.....
Data training 30	0.0077	0.5194	0.3038	0.0347	0.0093	0.5087	0.4444	0.0195	0.0073	0.4904	0.3269	0.0127	0.0056	0.4447	0.3125	0.019	0.005	0.4622	0.6769	0.0127

Tabel 4.31. menunjukkan arsitektur pada *backpropagation* seperti yang telah dimodelkan pada BAB III. Dimana neuron input untuk X1 – X4 adalah nilai untuk huruf “a”, X5 – X8 adalah nilai untuk huruf “e”, X9 – X12 adalah nilai untuk huruf “i”, X13 – X16 adalah nilai untuk huruf “o”, dan X17 – X20 adalah nilai untuk huruf “u”.

Kemudian nilai untuk X1 adalah nilai STFT yang ternormalisasi untuk huruf “a”, X2 adalah nilai amplitudo ternormalisasi untuk huruf “a”, X3 adalah nilai waktu ternormalisasi saat amplitudo tertinggi untuk huruf “a”, dan X4 adalah frekuensi ternormalisasi saat amplitudo tertinggi untuk huruf “a”. Demikian seterusnya untuk X5 – X20 dengan huruf vokal yang berbeda.

Keseluruhan nilai untuk X1 – X20 untuk 30 data pelatihan adalah nilai pada tabel 4.21 – tabel 4.30. Kemudian *backpropagation* akan melatih seluruh nilai tersebut dari perempuan yaitu data pelatihan 1 – 15, dan laki-laki untuk data pelatihan 16 – 30. Nilai-nilai tersebut disimpan kedalam sebuah tabel berformat (.mat) untuk dapat diakses oleh *Matlab*. Cara mengaksesnya adalah dengan membuat sintaks

```
traininput =
struct2cell(load('C:\Users\JosephMasarani\Documents\flashdisk
versi 7\Skripsi\matlab\dafuq\traininp5.mat'));

trainout =
struct2cell(load('C:\Users\JosephMasarani\Documents\flashdisk
versi 7\Skripsi\matlab\dafuq\trainout.mat'));

trainInp = traininput{1,1};
trainOut = trainout {1,1};
```

dimana *trainInp* adalah data latih yaitu data hasil ekstraksi ciri STFT, dan *trainOut* adalah data target.

Pelatihan ini bertujuan untuk mencari bobot optimal yang akan digunakan pada pengujian *backpropagation*, sehingga jaringan saraf tiruan dapat mengenali dan membedakan antara data perempuan dan laki-laki. Pelatihan antara perempuan dan laki-laki tidak dipisah, atau tidak mencari dua bobot optimal, yaitu bobot optimal untuk data perempuan, dan bobot optimal untuk data laki-laki. Hal ini dikarenakan urutan data untuk pengujian adalah urutan acak dari data perempuan dan laki-laki. Maka data yang dilatih adalah data dari perempuan sampai data laki-laki sehingga terbentuklah satu nilai bobot optimal yang bisa digunakan untuk kedua jenis kelamin suara tersebut.

Kecepatan pembelajaran data (*learning rate*) yang dipakai adalah 0.1. dan banyaknya perulangan (*iterasi*) adalah sebanyak 10000. Adapun penggunaan nilai *learning rate* 0.1 adalah agar jaringan dapat mencari *error* secara akurat, dan waktu yang dibutuhkan untuk pembelajaran menjadi lebih cepat. Jika dibandingkan dengan nilai *learning rate* 0.01, maka pencarian *error* akan lebih teliti, namun juga dibutuhkan waktu yang sangat lama untuk pembelajaran/pelatihan.

Jumlah iterasi juga menentukan nilai *error* yang dihasilkan. Semakin besar iterasi, maka semakin kecil juga nilai *error* yang dihasilkan. Kemudian jumlah neuron pada tiap *hidden layer* adalah sebanyak 150 *layer* pada *hidden layer* 1, dan 150 *layer* pada *hidden layer* 2. Jumlah tersebut didapatkan dari hasil *trial and error* dikarenakan nilai bobot awal yang digunakan adalah nilai acak yang berkisar antara -1 sampai 1, sehingga saat melakukan pelatihan lagi, nilai bobot

awal akan berubah dan tidak sama dengan nilai bobot awal pada pelatihan sebelumnya. Cara *trial and error* disini adalah untuk mencari bobot optimal yang ada pada setiap *hidden layer* untuk dapat mengenali jenis kelamin dari suara yang dimasukkan. Tabel 4.32. menunjukkan hasil percobaan untuk mencari nilai *error* terkecil dengan mengubah jumlah neuron pada *hidden layer* dan banyaknya iterasi.

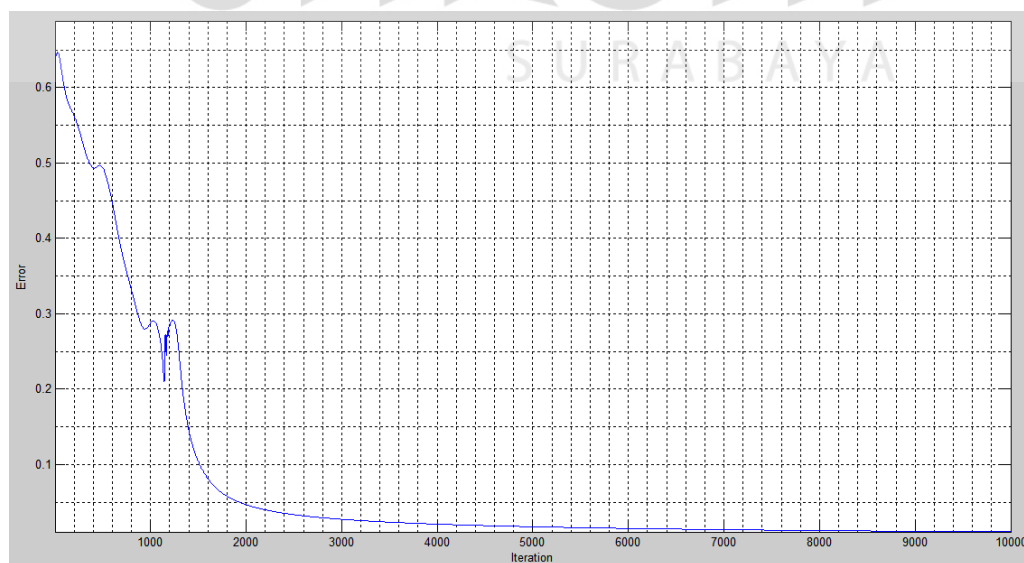
Tabel 4.28. Hasil *trial and error* untuk mencari nilai kuadrat galat error terkecil

iterasi	hidden 1	hidden 2	learn rate	threshold	mse	error data perempuan	error data laki-laki	ended
1000	20	20	0.1	0.01	0.4369	9	0	1000
1500	20	20	0.1	0.01	0.3075	4	0	1500
2000	20	20	0.1	0.01	0.1057	0	0	2000
3000	20	20	0.1	0.01	0.0294	0	0	3000
5000	20	20	0.1	0.01	0.0224	0	0	5000
2000	50	50	0.1	0.01	0.1070	0	0	2000
3000	50	50	0.1	0.01	0.0328	0	0	3000
5000	50	50	0.1	0.01	0.0166	0	0	5000
5000	100	100	0.1	0.01	0.0161	0	0	5000
10000	100	100	0.1	0.01	0.0111	0	0	10000
10000	100	100	0.01	0.01	0.0888	0	0	10000
3000	100	50	0.1	0.01	0.3758	5	2	3000
3000	50	100	0.1	0.01	0.3907	5	2	3000
3000	60	40	0.1	0.01	0.0339	0	0	3000
10000	150	150	0.1	0.01	0.0107	0	0	10000

Tabel 4.28. menunjukkan bahwa nilai mse yang didapatkan berpengaruh dari banyaknya iterasi dan jumlah neuron pada masing masing *hidden layer*. Seperti contoh saat banyaknya iterasi adalah 1000 dan jumlah neuron *hidden layer* 1 dan *hidden layer* 2 masing-masing adalah 20, nilai mse masih cukup besar yaitu 0.4369 dibandingkan dengan toleransi *error* yang digunakan adalah 0.01. sehingga jumlah *neuron hidden layer* dan banyaknya iterasi yang digunakan pada penelitin ini dengan adalah 150 *neuron hidden layer* 1, 150 *neuron hidden layer* 2, dan banyaknya iterasi adalah 10000 dengan nilai *learning rate* adalah 0.1, dan mse sebesar 0.0107.

Program akan terus melatih data, hingga *error* yang terjadi pada tiap iterasi kurang dari atau sama dengan toleransi *error* (threshold), atau selama banyaknya iterasi.

Gambar 4.19. adalah grafik dari *mean square error* pada percobaan iterasi sebanyak 10000, jumlah neuron pada masing-masing *hidden layer* adalah 150 dan nilai *learning rate* adalah 0.1.



Gambar 4.19. Grafik nilai *mean square error* pada tiap iterasi.

Pada gambar 4.19. dapat terlihat bahwa sistem akan berusaha meminimalkan nilai *mean square error* yang dihasilkan pada setiap iterasi. Error sudah mencapai dibawah 0.1 saat iterasi ke 1000 dan seterusnya. Saat error sudah sama dengan nilai *threshold*, dalam hal ini dibatasi sampai 0.01, atau jumlah iterasi sudah terpenuhi, maka bobot terakhir yang dipakai untuk pelatihan akan disimpan kedalam file (.mat) untuk nantinya akan digunakan dalam pengujian.

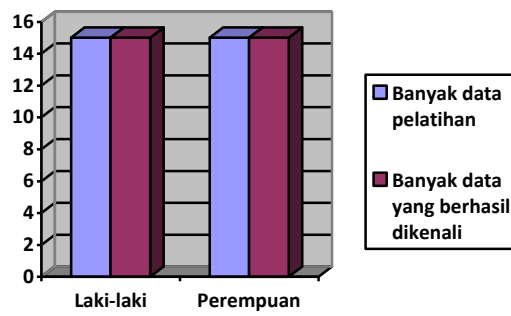
Setelah itu, bobot yang sudah tersimpan dalam file (.mat) akan digunakan untuk dijadikan bobot pada program pengujian. Program pengujian pada *backpropagation* hanya sampai alur maju (tidak menjalankan alur mundur), dengan nilai bobot yang sudah disimpan saat pelatihan.

Tabel 4.29 adalah hasil pengujian dari data *training* (pelatihan).

Tabel 4.29. Persentase keberhasilan pengenalan pada data *training*

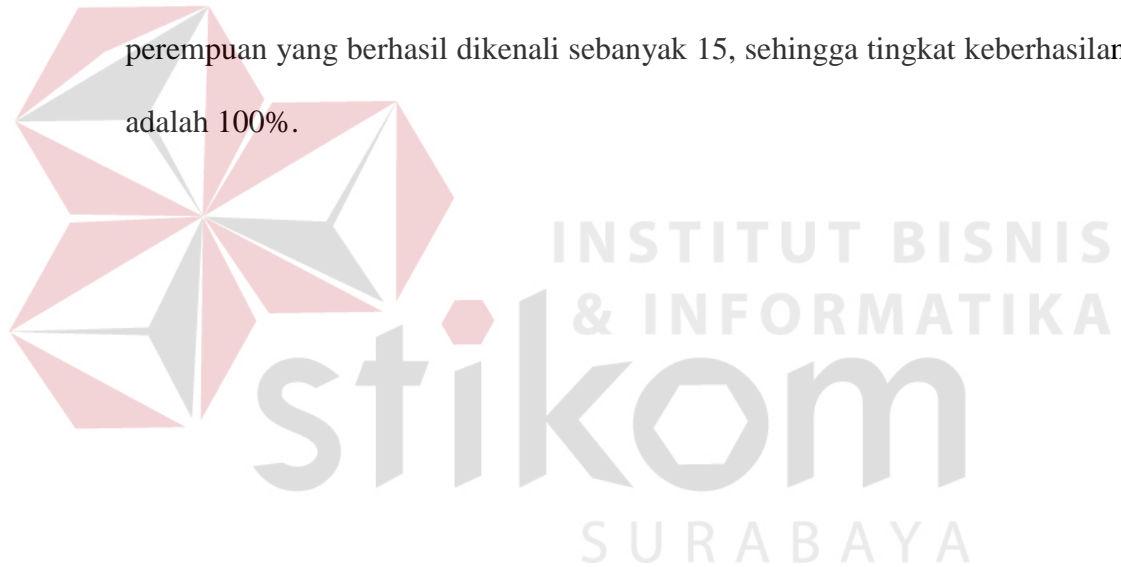
Jenis kelamin	Banyak data pelatihan (buah).	Data yang berhasil dikenali (buah).	Kesalahan pengenalan (buah)	Persentase Keberhasilan (%)
Laki-laki	15	15	0	100%
Perempuan	15	15	0	100%
Jumlah persentase total				100%

Pada pengujian menggunakan sampel data latih dalam tabel 4.29., jaringan saraf tiruan berhasil mengenali data latih perempuan dan laki-laki dengan tingkat keberhasilan adalah sebesar 100%.



Gambar 4.20. Grafik persentase keberhasilan pengenalan pada data training.

Gambar 4.20. menunjukkan persentase keberhasilan berdasarkan tabel 4.29, dimana data laki-laki yang berhasil dikenali adalah sebanyak 15, dan data perempuan yang berhasil dikenali sebanyak 15, sehingga tingkat keberhasilan adalah 100%.





```
Command Window
Perempuan, Data Training no. 1
Perempuan, Data Training no. 2
Perempuan, Data Training no. 3
Perempuan, Data Training no. 4
Perempuan, Data Training no. 5
Perempuan, Data Training no. 6
Perempuan, Data Training no. 7
Perempuan, Data Training no. 8
Perempuan, Data Training no. 9
Perempuan, Data Training no. 10
Perempuan, Data Training no. 11
Perempuan, Data Training no. 12
Perempuan, Data Training no. 13
Perempuan, Data Training no. 14
Perempuan, Data Training no. 15
Laki-laki, Data Training no. 16
Laki-laki, Data Training no. 17
Laki-laki, Data Training no. 18
Laki-laki, Data Training no. 19
Laki-laki, Data Training no. 20
Laki-laki, Data Training no. 21
Laki-laki, Data Training no. 22
Laki-laki, Data Training no. 23
Laki-laki, Data Training no. 24
Laki-laki, Data Training no. 25
Laki-laki, Data Training no. 26
Laki-laki, Data Training no. 27
Laki-laki, Data Training no. 28
Laki-laki, Data Training no. 29
Laki-laki, Data Training no. 30
Ended with 10000 iterations.

PPP =
0.0107

fx >>
```

Gambar 4.21. Hasil keluaran program pengujian data *training*.

Dapat dilihat pada grafik presentase dan tabel 4.29., bahwa sistem sudah dapat mengenali data. Terlihat pada gambar 4.21., bahwa sistem mengenali data training 1-15 adalah perempuan, dan 16-30 adalah laki-laki. Pelatihan data ini berlangsung selama 10000 iterasi dengan nilai MSE yaitu PPP adalah 0.0107.

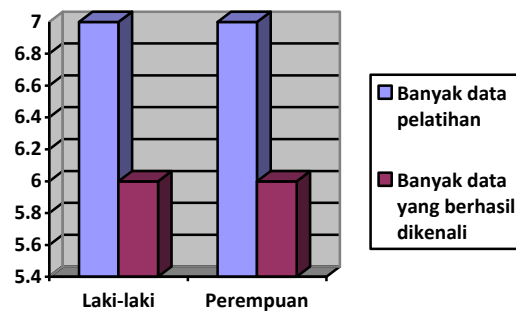
Dengan nilai bobot optimal yang sudah tersimpan tadi, maka akan diuji sebuah data baru dengan harapannya, data tersebut dapat langsung diidentifikasi jenis kelaminnya. Perlu diketahui bahwa pada pengujian, algoritma *backpropagation* cukup hanya dilakukan proses *feed forward*, sehingga nilai yang dikeluarkan langsung diidentifikasi.

Tabel 4.30 menunjukkan persentase keberhasilan pengenalan pada data baru

Tabel 4.30. Persentase keberhasilan pengenalan pada data baru

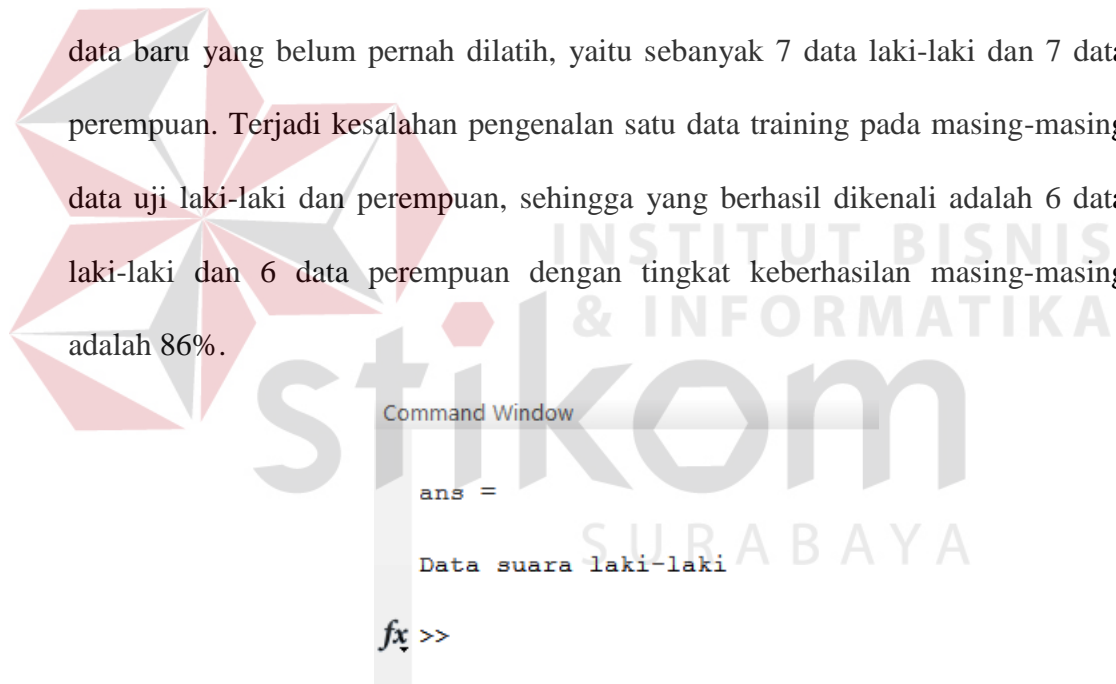
Jenis kelamin	Banyak data pelatihan (buah).	Data yang berhasil dikenali (buah).	Kesalahan pengenalan (buah)	Persentase Keberhasilan (%)
Laki-laki	7	6	1	86%
Perempuan	7	6	1	86%
Jumlah persentase total				86%

Saat pengujian pada data baru, sistem juga dapat mengidentifikasi suara laki-laki dan perempuan, namun terjadi kesalahan pada masing masing suara seperti terlihat pada tabel 4.30. Hal ini dapat terjadi karena ada kemiripan suara laki-laki pada perempuan saat perekaman suara, sehingga akan menghasilkan fitur-fitur suara hampir sama dengan laki-laki, maupun kemiripan suara perempuan pada suara laki-laki. Hal ini juga dapat dipengaruhi dari kurang banyaknya data pelatihan, sehingga untuk bisa mengenali, sistem perlu banyak mengenali banyak data untuk memahami berbagai macam pola. Dari tabel 4.30. dapat dibuat dalam bentuk grafik seperti pada gambar 4.22.



Gambar 4.22. Grafik persentase keberhasilan pengenalan pada salah satu data baru.

Gambar 4.22. dan tabel 4.30. menunjukkan bahwa presentase pengujian 14 data baru yang belum pernah dilatih, yaitu sebanyak 7 data laki-laki dan 7 data perempuan. Terjadi kesalahan pengenalan satu data training pada masing-masing data uji laki-laki dan perempuan, sehingga yang berhasil dikenali adalah 6 data laki-laki dan 6 data perempuan dengan tingkat keberhasilan masing-masing adalah 86%.



Gambar 4.23. Hasil keluaran program pengujian salah satu data baru laki-laki.

Gambar 4.23. merupakan keluaran program pengujian *backpropagation*, yaitu berupa status yang menunjukkan bahwa data yang diuji saat itu adalah data laki-laki. Dengan tampilan status tersebut maka program berhasil mengenali data suara laki-laki yang saat itu dijadikan sebagai data uji.