



**Identifikasi Akurasi Suara pada Audio Forensik dengan  
Metode Itakura Saito Distance dan Backpropagation Neural  
Network**

Ardy Wicaksono

18917107

*Tesis diajukan sebagai syarat untuk meraih gelar Magister Komputer*

*Konsentrasi Forensica Digital*

*Program Studi Informatika Program Magister*

*Fakultas Teknologi Industri*

*Universitas Islam Indonesia*

2020

**Lembar Pengesahan Pembimbing**

**Identifikasi Akurasi Suara pada Audio Forensik dengan Metode Itakura Saito  
Distance dan Backpropagation Neural Network**

Ardy Wicaksono

18917107



Yogyakarta, 07 Oktober 2020

المعتمد الاستاذ المساعد الدكتور  
Pembimbing 1

Pembimbing 2

(R.M Sisdarmanto Adinandra, S.T., M.Sc., Ph.D.)

(Dr. Yudi Prayudi, S.Si., M.Kom.)

## Lembar Pengesahan Penguji

### Identifikasi Akurasi Suara pada Audio *Forensic* dengan Metode Itakura Saito Distance dan Backpropagation Neural Network

Ardy Wicaksono

18917107

Yogyakarta, 07 Oktober 2020

Tim Penguji,

RM. Sisdamanto Adinandra, ST., M.Sc., Ph.D.

Ketua

Dr. Yudi Prayudi, S.Si., M.Kom.

Anggota I

Dr. Ir. Bambang Sugiantoro, S.Si., MT.

Anggota II

Mengetahui,

Ketua Program Studi Informatika Program Magister

Universitas Islam Indonesia



Izzati Muhammad, S.T., M.Sc., Ph.D.

## Abstrak

### Identifikasi Akurasi Suara pada Audio *Forensic* dengan Metode Itakura Saito Distance dan Backpropagation Neural Network

*Audio* merupakan salah satu barang bukti digital yang digunakan dalam *cybercrime*. Seringkali bahwa bukti *audio* ini membawa peran krusial untuk mengungkapkan adanya kasus kejahatan sehingga diperlukan proses analisis *audio forensic*. *Audio* tersebut berisi rekaman suara seseorang yang memiliki karakter dengan pengucapan kosakata yang berbeda-beda, pengucapan yang tidak jelas, dan memiliki banyak *noise*. Jika terjadi kasus tindakan kriminal dan ditemukan barang bukti *audio* berupa rekaman suara yang terduga sebagai pelaku, maka perlu adanya proses analisis untuk mencari kebenaran rekaman suara yang ditemukan identik dengan rekaman suara pelaku. Perlu adanya penanganan yang sesuai *Standart Operational Procedure* (SOP) *audio forensic*. Tahapan dalam melakukan *audio forensic* pada *Digital Forensic Analyst Team* Pusat Laboratorium Forensik (DFAT PUSLABFOR) terdiri dari 4 tahapan yakni *Acquisition*, *Audio Enhancement*, *Decoding*, dan *Voice Recognition*.

Dalam penelitian ini akan dibahas mengenai analisis *audio* menggunakan metode *speech processing* yaitu *Itakura Saito Distance* dan metode jaringan syaraf tiruan yaitu *Backpropagation Neural Network* dengan tujuan memperkuat hasil akurasi identik suatu barang bukti rekaman suara. Jika metode ini dikaloborasikan akan memperkuat tingkat akurasi dan *argument* yang diperoleh dari proses analisa, khususnya dalam penanganan *audio forensic*. Akurasi itu sendiri diukur dari nilai kedekatan frekuensi atau *spectrum* antara rekaman suara asli dengan rekaman suara pembanding. Nilai presentase tinggi akurasi dari metode *Itakura Saito Distance* masih menjadi perdebatan. Asumsi tersebut disampaikan oleh seorang saksi ahli dalam kasus tindak pidana korupsi di Indonesia tentang barang bukti rekaman suara, dimana metode ini banyak digunakan dalam rekayasa *audio* yaitu *background* musik ataupun konser, tetapi tidak mengarah untuk membandingkan dua suara.

Pengujian dilakukan pada 4 rekaman suara asli (*unknown*) dan 4 rekaman suara pembanding (*known*) menggunakan lebih dari 20 kosakata dengan proses pertama menganalisa metode *Itakura Saito Distance*. Analisa menggunakan batas panjang waktu (*frame*) sebesar 100000 dari keseluruhan panjang waktu dari setiap suara itu sendiri. Hasil dari metode tersebut merupakan nilai *formant* yang selanjutnya dianalisa kembali dengan

metode *Backpropagation Neural Network* menggunakan ketepatan parameter yang dimasukkan. Penggabungan dari metode tersebut menunjukkan akurasi tertinggi yang identik lebih dari 95%.

**Kata kunci**

*Itakura Saito Distance, Backpropagation Neural Network, Audio Forensik, Akurasi Suara*



## **Abstract**

### **Identification of Sound Accuracy in Audio Forensic by the Itakura Saito Distance and Backpropagation Neural Network Methods**

Audio is one of the digital evidence used in cybercrime. Often this audio evidence plays a crucial role in revealing a crime case, so an audio forensic analysis process is required. The audio contains a recorded voice of a person who has a character with different vocabulary pronounced, unclear pronunciation, and has a lot of noise. If a criminal act occurs and audio evidence is found in the form of a voice recording that is suspected of being the perpetrator, it is necessary to have an analysis process to find the truth of the voice recording that is found to be identical with the perpetrator's voice recording. There needs to be an appropriate handling of the audio forensic Standard Operational Procedure (SOP). The stages in conducting audio forensics on the Forensic Laboratory Center Digital Forensic Analyst Team (DFAT PUSLABFOR) consist of 4 stages, namely Acquisition, Audio Enhancement, Decoding, and Voice Recognition.

This research will discuss about audio analysis using speech processing method, namely Itakura Saito Distance and artificial neural network method, namely Backpropagation Neural Network with the aim of strengthening the identical accuracy of a sound recording evidence. If this method is used, it will strengthen the level of accuracy and arguments obtained from the analysis process, especially in handling audio forensics. Accuracy itself is measured from the frequency proximity value or spectrum between the original sound recording and the comparator sound recording. The high percentage accuracy of the Itakura Saito Distance method is still a matter of debate. This assumption was conveyed by an expert witness in a criminal case of corruption in Indonesia regarding sound recorded evidence, where this method is widely used in audio engineering, namely music or concert backgrounds, but does not lead to comparing two voices.

Tests were carried out on 4 unknown voice recordings and 4 known voice recordings using more than 20 vocabularies with the first process of analyzing the Itakura Saito Distance method. The analysis uses a time limit (frame) of 100 000 of the total length of time of each sound itself. The result of this method is the formant value which is then re-analyzed using the Backpropagation Neural Network method using the accuracy of the parameters entered.

The combination of these methods shows the highest accuracy which is identical to more than 95%.

**Keyword**

Itakura Saito Distance, Backpropagation Neural Network, Audio Forensic, Sound Accuracy



## **Pernyataan Keaslian Tulisan**

Dengan ini saya menyatakan bahwa tesis ini merupakan tulisan asli dari penulis, dan tidak berisi material yang telah diterbitkan sebelumnya atau tulisan dari penulis lain terkecuali referensi atas material tersebut telah disebutkan dalam tesis. Apabila ada kontribusi dari penulis lain dalam tesis ini, maka penulis lain tersebut secara eksplisit telah disebutkan dalam tesis ini.

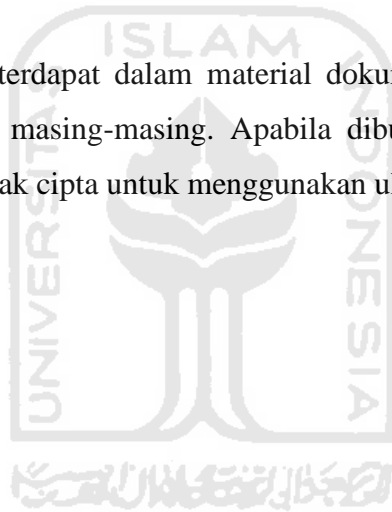
Dengan ini saya juga menyatakan bahwa segala kontribusi dari pihak lain terhadap tesis ini, termasuk bantuan analisis statistik, desain survei, analisis data, prosedur teknis yang bersifat signifikan, dan segala bentuk aktivitas penelitian yang dipergunakan atau dilaporkan dalam tesis ini telah secara eksplisit disebutkan dalam tesis ini.

Segala bentuk hak cipta yang terdapat dalam material dokumen tesis ini berada dalam kepemilikan pemilik hak cipta masing-masing. Apabila dibutuhkan, penulis juga telah mendapatkan izin dari pemilik hak cipta untuk menggunakan ulang materialnya dalam tesis ini.

Yogyakarta, 07 Oktober 2020



(Ardy Wicaksono, S.Kom.)



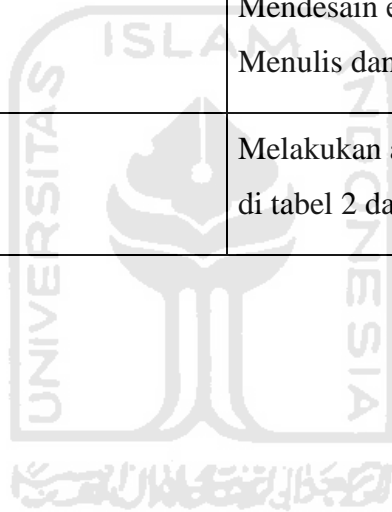


## Daftar Publikasi

Paper yang dihasilkan:

(Ardy Wicaksono, Sisdarmanto Adinandra, Yudi Prayudi, 2020), Penggabungan Metode Itakura Saito Distance dan Backpropagation Neural Network untuk Peningkatan Akurasi Suara pada Audio Forensik, *JUITA J. Inform.*, Vol. 8, No. 2, Edisi November 2020).

Kontributor	Jenis Kontribusi
Ardy Wicaksono	Mendesain eksperimen (70%) Menulis <i>paper</i> (80%)
Sisdarmanto Adinandra	Mendesain eksperimen (30%) Menulis dan mengedit <i>paper</i> (20%)
Yudi Prayudi	Melakukan analisis statistik dari data di tabel 2 dan tabel 3



## Halaman Persembahan

*Alhamdulillah Robbil 'Alamin.* Segala puji dan syukur atas kehadiran Allah Subhana Wa Ta'ala yang telah memberikan rahmat, ridho, dan karunia-Nya kepada saya. Shalawat serta salam kepada Nabi Muhammad Shallallahu 'Alaihi Wasallam, sebagai pembawa risalah Allah terakhir dan penyempurna seluruh risalah-Nya. Tesis ini kupersembahkan kepada:

1. **Allah SWT** yang telah memberiku nikmat iman dan islam, serta sang guru besarku **Nabi Agung Muhammad SAW** yang mengajarkanku ilmu akan arti kehidupan.
2. **Kedua Orang tua** tersayang yang selalu mengiringi doa, motivasi, serta nasehat dalam hidupku. Tiada kata yang dapat ku tulis untuk menggambarkan segala pengorbanan dan kasih sayang kalian. Namun hanya doa yang dapat ku persembahkan semoga kasih sayang dan rahmat Allah SWT senantiasa tercurahkan.
3. **Dosen dan Seluruh pengurus Akademik MI UII** yang berjasa serta bersedia memberikan waktu dan ilmu pengetahuan selama menempuh masa studi magister.
4. Untuk orang ku sayang, **Wan Tawang Oktaviana Nur Halimah** yang telah memberiku semangat, dorongan positif dan senantiasa selalu menemaniku dalam menyusun laporan tesis ini.
5. **Teman, Sahabat Kos Ronald Glagahsari dan MI UII 2018 khususnya Konsentrasi Forensika Digital** yang selalu memberikan semangat, motivasi, dan pengertian menempuh masa studi magister.
6. Teman-teman yang tidak dapat penulis sebutkan satu-persatu yang ikut mendukung penulis dalam penyusunan Laporan Tesis ini, maupun dalam menempuh masa studi magister.

## Kata Pengantar

*Assalamu'alaikum Wr. Wb.*

Puji syukur penulis sampaikan kehadiran Allah SWT., Tuhan Yang Maha Pemurah lagi Maha Penyayang. Berkat rahmat, hidayah, dan inayah-Nya, akhirnya penulis dapat menyelesaikan laporan tesis yang berjudul “Identifikasi Akurasi Suara pada Audio *Forensic* dengan Metode *Itakura Saito Distance* dan *Backpropagation Neural Network*”.

Tesis ini merupakan syarat wajib yang harus ditempuh dalam mencapai Magister Strata-2 pada Program Studi Magister Informatika. Penulisan laporan tesis ini terselesaikan karena bantuan dari berbagai pihak. Untuk itu, penulis menyampaikan terima kasih kepada semua pihak yang telah membantu penulis dalam menyelesaikan laporan tesis ini.

1. Bapak Prof. Fathul Wahid, S.T., M.Sc., Ph.D., Selaku Rektor Universitas Islam Indonesia.
2. Bapak Hari Purnomo, Prof., Dr., Ir., M.T., Selaku Dekan Fakultas Teknologi Industri Universitas Islam Indonesia.
3. Ibu Izzati Muhimmah, S.T., M.Sc., Ph.D., Selaku Ketua Program Studi Magister Informatika Fakultas Teknologi Industri Universitas Islam Indonesia.
4. Bapak Dr. R.M. Sisdarmanto Adinandra, S.T., M.Sc., Selaku Dosen Pembimbing 1 yang berkenan membimbing dan mengarahkan penulis, sehingga Penelitian Tesis ini dapat selesai.
5. Bapak Dr. Yudi Prayudi, S. Si., M. Kom., Selaku Dosen Pembimbing 2 yang berkenan membimbing dan mengarahkan penulis, sehingga Penelitian Tesis ini dapat selesai.
6. Semua pihak yang tidak dapat penulis sebutkan satu persatu, baik secara langsung maupun tidak langsung membantu dalam penulisan tesis ini.

Semoga atas bantuan dan kerjasama yang telah diberikan menjadi amal baik dan mendapat balasan dari Allah SWT. Penulis menyadari bahwa penelitian ini jauh dari sempurna, untuk itu diperlukan saran dan masukan demi sempurnanya penyusunan laporan tesis ini. Akhir kata, penulis berharap semoga laporan tesis ini dapat bermanfaat bagi semua yang membutuhkan.

*Wassalamu'alaikum Wr. Wb.*

Yogyakarta, September 2020

(Ardy Wicaksono)

## Daftar Isi

Lembar Pengesahan Pembimbing .....	i
Lembar Pengesahan Penguji.....	ii
Abstrak .....	iii
Abstract.....	v
Pernyataan Keaslian Tulisan .....	vii
Daftar Publikasi .....	viii
Halaman Persembahan .....	ix
Kata Pengantar.....	x
Daftar Isi.....	xi
Daftar Tabel.....	xiii
Daftar Gambar .....	xiv
Glosarium .....	xvi
BAB 1 Pendahuluan .....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Batasan Masalah .....	3
1.4. Tujuan Penelitian .....	3
1.5. Manfaat Penelitian .....	4
1.6. Review Penelitian .....	4
1.7. Metode Penelitian .....	7
1.8. Sistematika Penelitian.....	8
BAB 2 Tinjauan Pustaka .....	9
2.1. Audio .....	9
2.2. <i>Audio Forensic</i> .....	12
2.3. <i>Lowpass Filter</i> .....	13

2.4.	Fast Faurier Transform (FFT).....	15
2.5.	Akurasi Suara.....	16
2.6.	Itakura Saito Distance.....	16
2.7.	Jaringan Syaraf Tiruan.....	18
2.8.	Backpropagation.....	21
BAB 3 Metodologi.....		26
3.1.	Tinjauan Pustaka.....	27
3.2.	Teknik Pengambilan Sampel Data.....	27
3.3.	Implementasi Metode.....	31
3.4.	Analisis Hasil Data.....	34
3.5.	Laporan.....	35
BAB 4 Hasil dan Pembahasan.....		36
4.1.	Analisis Kebutuhan.....	36
4.2.	Skenario Kasus.....	36
4.3.	Analisis Metode.....	37
4.3.1.	Itakura-Saito Distance.....	38
4.3.2.	<i>Backpropagation Neural Network</i> .....	44
4.4.	Hasil Analisis Penggunaan Metode Itakura-Saito Distance dengan Metode Backpropagation Neural Network.....	48
BAB 5 Kesimpulan dan Saran.....		53
5.1.	Kesimpulan.....	53
5.2.	Saran.....	53
Daftar Pustaka.....		55
LAMPIRAN.....		58

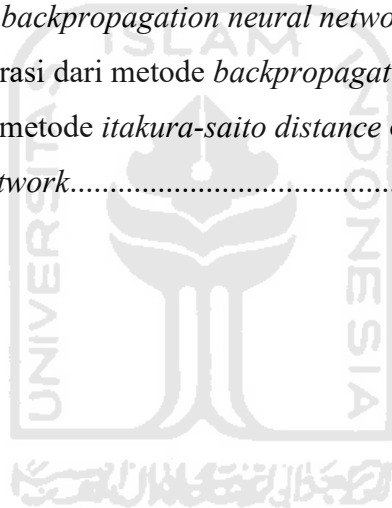
## Daftar Tabel

Tabel 1. 1 Tabulasi Studi Pustaka dan Pengembangan yang dilakukan.....	6
Tabel 1. 2 Lanjutan.....	7
Tabel 3. 1 Analisis Implementasi Metode.....	34
Tabel 3. 2 Hasil evaluasi metode.....	35
Tabel 3. 3 Karakter hasil analisis metode.....	35
Tabel 4. 1 Penentuan data analisis.....	37
Tabel 4. 2 Hasil evaluasi perbandingan nilai itakura-saito distance pada Rekaman A1 dan Rekaman P1, P2, P3, P4.....	42
Tabel 4. 3 Hasil evaluasi perbandingan nilai itakura-saito distance pada Rekaman A2 dan Rekaman P1, P2, P3, P4.....	43
Tabel 4. 4 Hasil evaluasi perbandingan nilai <i>itakura-saito distance</i> pada Rekaman A3 dan Rekaman P1, P2, P3, P4.....	43
Tabel 4. 5 Hasil evaluasi perbandingan nilai itakura-saito distance pada Rekaman A4 dan Rekaman P1, P2, P3, P4.....	43
Tabel 4. 6 Hasil Akurasi identifikasi rekaman suara dengan metode <i>itakura-saito distance</i> dan <i>backpropagation neural network</i> .....	49
Tabel 4. 7 Perbedaan metode Konvensional dengan metode <i>Itakura Saito Distance</i> dan metode <i>Backpropagation Neural Network</i> .....	49
Tabel 4. 8 Lanjutan.....	50
Tabel 4. 9 Jenis analisa dan karakter dari masing-masing metode.....	51
Tabel 4. 10 Perbandingan hasil akurasi identifikasi rekaman suara <i>unknown</i> dan <i>known</i> ..	52

## Daftar Gambar

Gambar 2. 1 Diagram <i>pitch</i> terhadap waktu berubah secara konstan .....	10
Gambar 2. 2 Diagram masing-masing Formant F1, F2, F3, F4 dan F5. ....	11
Gambar 2. 3 Representasi spectrogram spektral dengan tingkatan energinya. ....	12
Gambar 2. 4 Jenis-jenis <i>filter</i> .....	14
Gambar 2. 5 <i>low-pass filter specifications</i> .....	15
Gambar 2. 6 Diagram blok LPC untuk <i>speech recognition</i> .....	17
Gambar 2. 7 Model sederhana JST.....	19
Gambar 2. 8 Arsitektur <i>single layer network</i> .....	20
Gambar 2. 9 Arsitektur <i>multiplayer network</i> .....	20
Gambar 3. 1 Alur metode penelitian .....	26
Gambar 3. 2 Ilustrasi pengambilan sampel rekaman suara .....	27
Gambar 3. 3 Prosedur penanganan barang bukti digital rekaman suara .....	28
Gambar 3. 4 Alur Proses Akusisi Barang Bukti Digital.....	29
Gambar 3. 5 Ilustrasi proses <i>audio enhancement</i> .....	29
Gambar 3. 6 Tampilan <i>effects DeReverb</i> .....	29
Gambar 3. 7 Perbedaan hasil rekaman suara pada proses penghilangan <i>noise</i> .....	30
Gambar 3. 8 Tampilan halaman kerja oTranscribe .....	30
Gambar 3. 9 Hasil transkrip percakapan dari <i>audio sample</i> barang bukti.....	31
Gambar 3. 10 Tahap implementasi metode <i>Itakura-Saito Distance</i> .....	32
Gambar 3. 11 Implementasi metode <i>Backpropagation Neural Network</i> .....	33
Gambar 3. 12 Implementasi proses <i>training</i> metode <i>Backpropagation Neural Network</i> ...	34
Gambar 4. 1 Ilustrasi Skenario Kasus .....	37
Gambar 4. 2 Sinyal analog Rekaman A1 dan P2 .....	38
Gambar 4. 3 Proses melakukan low-pass filter .....	38
Gambar 4. 4 Hasil dari memperbaiki kualitas suara dengan <i>low-pass filter</i> .....	39
Gambar 4. 5 Perhitungan LPC.....	39
Gambar 4. 6 Perintah pemanggilan <i>function</i> LPC.....	40
Gambar 4. 7 Hasil nilai LPC $y_1$ dari rekaman suara A1 .....	40
Gambar 4. 8 Perintah mengkalkulasi LPC dengan <i>function lpcar2ra</i> .....	40
Gambar 4. 9 Menghitung nilai jarak pada metode <i>Itakura-Saito Distance</i> .....	41

Gambar 4. 10 Perintah menjalankan <i>function Itakura-Saito Distance</i> .....	41
Gambar 4. 11 Hasil nilai <i>spectrum (formant)</i> dari proses <i>itakura-saito distance</i> .....	41
Gambar 4. 12 Perintah pencarian akurasi suara identik .....	42
Gambar 4. 13 Hasil Visualisasi <i>Itakura-Saito Distance</i> .....	42
Gambar 4. 14 Save hasil <i>itakura-saito distance</i> dengan ekstensi ".mat" .....	44
Gambar 4. 15 Proses <i>load data</i> dan normalisasi .....	44
Gambar 4. 16 Proses <i>invers</i> matriks .....	45
Gambar 4. 17 <i>Input</i> parameter <i>training</i> .....	45
Gambar 4. 18 Arsitektur <i>newff</i> .....	46
Gambar 4. 19 Perintah proses <i>Backpropagation Neural Network</i> .....	46
Gambar 4. 20 Proses <i>training Backpropagation Neural Network</i> .....	46
Gambar 4. 21 Informasi detail variabel "net" .....	47
Gambar 4. 22 Proses pengujian <i>backpropagation neural network</i> .....	47
Gambar 4. 23 Analisis nilai akurasi dari metode <i>backpropagation neural network</i> .....	48
Gambar 4. 24 Hasil Visualisasi metode <i>itakura-saito distance</i> dengan metode <i>backpropagation neural network</i> .....	48





## Glosarium

CoC	- <i>Chain of Custody</i>
SOP	- Standar Operasional Prosedur
BPNN	- <i>Backpropagation Neural Network</i>
DFAT	- <i>Digital Forensic Analyst Team</i>
ACPO	- <i>Association of Chief Police Officers</i>
PUSLABFOR	- Pusat Laboratorium Forensik
LP	- <i>Linear Prediction</i>
LPC	- <i>Linear Prediction Coding</i>
SSE	- <i>Sum of Squared Error</i>
JST	- Jaringan Syaraf Tiruan
MSE	- <i>Mean Square Error</i>



# BAB 1

## Pendahuluan

### 1.1. Latar Belakang

*Audio* merupakan salah satu jenis barang bukti digital. Barang bukti yang diajukan dipersidangan harus terjamin untuk mencegah gangguan bukti, khususnya barang bukti digital (Prayudi, dkk., 2014). Demi memenuhi keuntungan beberapa orang dipengadilan rekaman percakapan dari barang bukti *audio* sering sekali dilakukan manipulasi atau pengeditan (Imran, dkk 2017). Dalam beberapa kasus, *audio* menjadi barang bukti yang sering menjadi perdebatan khususnya dalam mengukur dan mendeteksi keakuratan suara. Sehingga dengan adanya hal tersebut diperlukan pendukung keputusan dalam memverifikasi *audio forensic* (Renza, dkk, 2018).

Kedudukan barang bukti digital sangat penting dalam pembuktian suatu kasus, sehingga diperlukan saksi ahli yang kompeten khususnya dibidang keilmuan *audio forensic* (Handoko, 2017). Seorang saksi ahli harus mampu memaparkan analisis yang relevan menyeluruh dan otentik dalam melakukan serangkaian pemeriksaan. Membuktikan suatu peristiwa khususnya barang bukti *audio* harus sesuai dengan ketentuan, karena menyangkut nasib dari korban maupun tersangka (Andri & Tawang, 2018). Serangkaian pemeriksaan harus mengikuti Undang-Undang, Standar Operasional Prosedur (SOP), dan sesuai dengan ketentuan tahapan *audio forensic* pada DFAT PUSLABFOR (Putri & Sunarno, 2014).

Kemampuan untuk melakukan analisis *audio forensic* tergantung pada ketersediaan rekaman *audio* yang diperoleh. Rekaman *audio* berisi wawancara, penyadapan rahasia, serta merekam interogasi (Maher, 2018). Rekaman *audio* yang ditemukan memiliki karakteristik berbeda-beda pada setiap orangnya dan rekaman yang diperoleh kualitasnya belum tentu baik. Rekaman suara dengan merendahkan *pitch* cenderung mudah diidentifikasi, dibandingkan rekaman meninggikan *pitch* (Subki, dkk., 2018). Jika terjadi kasus tindakan kriminal dan ditemukan barang bukti *audio* berupa rekaman suara yang terduga sebagai pelaku, maka perlu adanya proses analisis untuk mencari kebenaran rekaman suara yang ditemukan identik dengan rekaman suara pelaku. Analisis barang bukti digital yang berupa audio dapat dilakukan dengan metode berbasis *autorecorder* atau *automatic statistic* yaitu *Itakura Saito Distance*. Metode ini menggunakan jarak antara *spectrum* suara asli dan suara perkiraan sebagai indikator akurasi barang bukti *audio* (Nozaki & Nakamoto, 2017).

*Autoencoder* pada suara yang bergema merekonstruksi pengenalan ucapan dengan cara *spectrum* jangka pendek dan jangka panjang diuji secara matematis, sehingga memperoleh akurasi dari rekaman suara yang diperoleh (Ishii, dkk., 2013).

Metode *Itakura Saito Distance* merupakan metode *speech processing* yang dapat menggunakan *linier prediction coding* dalam mengukur perbedaan *spectrum* suara asli dan *spectrum* suara perkiraan. Dari proses perhitungan LPC dengan metode *Itakura Saito Distance* dapat diperoleh nilai akurasi yang identik apabila semakin dekat jarak antara dua *spectrum*, maka semakin mirip perbandingan suara tersebut (Irawan, 2019). Akurasi itu sendiri diukur dari nilai kedekatan frekuensi atau *spectrum* antara rekaman suara asli dengan rekaman suara pembanding. Nilai presentase tinggi akurasi dari metode *Itakura Saito Distance* masih menjadi perdebatan. Asumsi tersebut disampaikan oleh seorang saksi ahli dalam kasus tindak pidana korupsi di Indonesia tentang barang bukti rekaman suara, dimana metode ini banyak digunakan dalam rekayasa *audio* yaitu *background* musik ataupun konser, tetapi tidak mengarah untuk membandingkan dua suara (Hukumonline.com, 2019). Analisis yang sudah dilakukan dari proses identifikasi rekaman suara asli dengan rekaman suara pembanding dengan metode *Itakura Saito Distance* dan Metode Konvensional (*pitch, formant, spectrogram*) memperoleh nilai presentase akurasi identik tertinggi sebesar 90,83% (Irawan, 2019).

Metode BPNN merupakan algoritma yang efektif dalam memberikan solusi penurunan gradien untuk meminimalkan kuadrat *error input*. Metode BPNN mempunyai tiga tahapan utama yaitu tahap perambatan maju, tahap perambatan balik, dan tahap perubahan bobot dan bias (Huang, dkk., 2015). Gradien dari fungsi dihitung sehubungan dengan bobot dan dalam bias jaringan yang meminimalkan fungsi tersebut (Nozaki & Nakamoto, 2017). Tujuan utama metode BPNN adalah mendapatkan keseimbangan antara pengenalan pola pelatihan secara benar dan respon yang baik untuk pola lain yang sejenis. Umumnya pola data dibagi menjadi dua bagian, yaitu pola data yang dipakai sebagai pelatihan dan data yang dipakai untuk pengujian. Perubahan bobot dilakukan berdasarkan pola pelatihan. Selama pelatihan dalam *epoch* tertentu kesalahan yang terjadi dihitung berdasarkan data pelatihan dan pengujian (Hilmi, 2014). Jika nilai *error* menurun maka akan melakukan proses pelatihan selanjutnya. Akan tetapi jika kesalahannya sudah meningkat, pelatihan tidak diteruskan lagi (Sutojo, dkk., 2011).

Kasus tindakan kriminal yang melibatkan barang bukti rekaman suara masih menjadi perdebatan tersendiri dari bentuk karakter suara yang dihasilkan dari rekaman *audio*,

maupun algoritma yang belum pasti mengenai perbandingan identik atau tidak rekaman suara tersebut. Metode yang sudah disebutkan diatas memiliki karakteristik dan parameter yang berbeda-beda, akan tetapi memiliki tujuan yang sama. Metode *Itakura Saito Distance* diasumsikan mempunyai nilai identik akurasi yang lebih baik dalam membandingkan suara *unknown* dan *known*, kemudian hasil dari metode tersebut diekstraksi kembali atau diperkuat dengan metode BPNN, sehingga akan mendapatkan nilai yang relevan dalam menganalisis bukti digital *file* rekaman *audio* sebagai barang bukti digital sesuai dengan ketentuan *Chain of Custody* (CoC). Perolehan akurasi hasil identik antara dua suara dari metode *Itakura Saito Distance* masih ada kemungkinan ditingkatkan kembali nilai akurasinya dengan menggunakan metode *Backpropagation Neural Network*, sekaligus memperkuat *argument* penanganan dalam proses *audio forensic*.

## 1.2. Rumusan Masalah

Adapun rumusan masalah pada penelitian ini adalah:

1. Bagaimana melakukan kombinasi metode *Itakura-Saito Distance* dengan metode *Backpropagation Neural Network* untuk meningkatkan hasil analisis akurasi perbandingan suara?
2. Bagaimana kinerja penggabungan metode *Itakura-Saito Distance* dengan metode *Backpropagation Neural Network* untuk meningkatkan hasil analisis akurasi perbandingan suara?

## 1.3. Batasan Masalah

Batasan masalah pada penelitian *Audio Forensic* ini sebagai berikut:

1. File yang diuji dan dianalisis adalah file audio dengan ekstensi *.wav*;
2. Terdapat dua (2) jenis sampel rekaman suara yang digunakan untuk perbandingan yaitu rekaman suara bukti digital asli dan suara rekaman pembanding;
3. Parameter analisis statistik yang digunakan adalah nilai *formant*.

## 1.4. Tujuan Penelitian

Tujuan penelitian pada *Audio Forensic* ini sebagai berikut:

1. Melakukan kombinasi metode *Itakura-Saito Distance* dan metode *Backpropagation Neural Network* untuk mendapatkan akurasi identifikasi suara pada *audio forensic*.

2. Melakukan ujicoba terhadap metode *Itakura-Saito Distance* dengan metode *Backpropagation Neural Network* yang telah dibuat dengan kasus perbandingan rekaman suara asli dan suara pembandingan.

### 1.5. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan kontribusi dalam kehidupan manusia dan dapat diterapkan di dunia nyata, khususnya membantu pembuktian barang bukti digital dipersidangan. Adapun manfaat penelitian ini antara lain:

1. Mengetahui berbagai metode *audio forensic* yang relevan untuk melakukan perbandingan dan analisis kemiripan suara pada *file* barang bukti rekaman *audio*.
2. Membantu penegak hukum untuk mengetahui keaslian pemilik suara dari rekaman asli yang diperoleh dari barang bukti digital dengan nilai keakuratan bisa dipertanggungjawabkan.

### 1.6. Review Penelitian

Teknologi *audio* memiliki peranan yang sangat penting dan menunjukkan dampak yang besar seperti penelitian yang dilakukan oleh (Nozaki & Nakamoto, 2017) dengan menggunakan metode *Itakura-Saito Distance* berbasis *autoencoder* atau *automatic statistic*. Puncak terbesar dalam data *sheet* biasanya lebih ditekankan dari puncak terkecil dan *autoencoder* banyak digunakan untuk pengurangan dimensi *nonlinear* atau bentuk khusus. Melakukan pengurangan dimensi dari spektrum massa menggunakan metode jaringan saraf tiruan untuk mendapatkan terkompresi dan representasi distribusi. *Autoencoder* yang menggunakan jarak sebagai fungsi biaya seperti *square error* dan lintas entropi. Setelah menemukan nilai terkecil maka akan dianalisis secara *automatic statistic* dengan tujuan mendapatkan hasil keluaran yang diinginkan.

Penerapan *autoencoder denoising* pada pengenalan suara yang bergema sebagai ujung yang kuat seperti penelitian yang dilakukan (Ishii, dkk, 2013) untuk merekonstruksi ucapan yang spektrum dari input yang berisik. Untuk menangkap efek konteks bunyi suara memerlukan bingkai *spectral* pendek yang disatukan untuk membentuk *vector input* tunggal. Kombinasi *spectra* jangka pendek maupun panjang diselidiki untuk menangani *impuls* panjang suara dan menjaga resolusi waktu yang diperlukan. Penelitian ini menggunakan *autoencoder denoising* berbasis *front-end* menggunakan spektrum *shortwindowed* yang memberikan hasil lebih baik dari pada metode konvensional. Akurasi pengenalan suara dengan metode tersebut menggunakan *spectra* jangka pendek dan jangka panjang untuk uji

kondisi terbuka dari *dataset*. Kemudian sebagai percobaan menggunakan pengenalan ucapan kosakata besar dan efektivitas metode.

Meminimalkan *Itakura-Saito Distance* ke metode *apriori* dengan daya spektrum dibawah asumsi *gaussion* dapat mengurangi kesalahan prediksi dalam masalah pencocokan *audio*. Terdapat dua *spectra* yaitu *spectra entropy* maksimum dan *collegram*, pada *spectra entropy* dengan memberikan nilai *spectra* kepada *apriori* sehingga menentukan seberapa dekat perkiraan urutan. Sehingga pada penelitian yang dilakukan (Enqvist & Karlsson, 2008) dapat ditarik kesimpulan pendekatan untuk mengidentifikasi subjek kekuatan *spectra* saat menggunakan informasi *apriori* tergantung pada penemuan *spectra* yang paling dekat dengan *spectra* yang diberikan *Itakura-Saito Distance*.

Penelitian tentang *audio forensic* juga dilakukan oleh (Putri & Sunarno, 2014) yang menganalisis rekaman suara menggunakan teknik *audio forensic* untuk keperluan penanganan barang bukti digital. Penelitian ini menekankan bahwa dalam suatu persidangan, barang bukti dan alat bukti rekaman suara yang ditemukan dapat mengungkap identik atau tidaknya suara rekaman asli suara dari terduga pelaku. Penggunaan algoritma analistik statistik dapat menemukan suara identik atau tidak dari suara asli dengan suara pembanding secara manual dengan menggunakan parameter *pitch*, *LR formant bandwidth*, *anova statistic*, analisis sebaran grafis, dan analisis *spectrogram*.

Membandingkan metode *Itakura-Saito Distance* dan metode konvensional dalam *audio forensic* juga dilakukan oleh (Irawan, 2019) yang mekanismenya dalam proses analisis perbandingan rekaman barang bukti suara, dimana metode konvensional mampu menganalisis kriteria kosakata yang jelas dan kualitas suara yang baik karena memiliki beberapa analisis seperti adalah *pitch*, *formant*, *likelihood*, *graphical distribution*, dan *spectrogram*. Sedangkan metode *Itakura-Saito Distance* mampu menganalisis dalam bentuk kalimat dengan kualitas suara yang kurang baik, karena hanya menganalisis jarak nilai *formant* dan memvisualisasi kedalam bentuk grafik. Hal tersebut mengakibatkan analisis metode konvensional membutuhkan waktu yang lebih lama dibandingkan metode *Itakura-Saito Distance*, akan tetapi metode konvensional mampu menganalisis rekaman suara yang lebih detail dibandingkan dengan metode *Itakura-Saito Distance*. Implementasi metode konvensional dalam menganalisis perbandingan suara memerlukan *tools* tambahan untuk menghilangkan *noice*, sedangkan metode *Itakura-Saito Distance* mem-filter *noice* atau menghilangkan *frame* yang tidak dibutuhkan secara otomatis atau *Linear Predictive Coding* (LPC).

Penelitian identifikasi suara yang dilakukan (Hilmi, 2014) menerapkan kombinasi antara *digital signal processing* dengan *fast fourrier transform* dan pengenalan pola suara dengan menggunakan metode jaringan syaraf tiruan (*neural network*). Tempat pengambilan suara dan jarak rekaman suara sangat mempengaruhi hasil pengujian. Penelitian mampu mengidentifikasi suara dengan FFT dan metode *backpropagation* dengan hasil tertinggi yaitu 93% sedangkan variasi jaringan memakai 3 *hidden layer*, *learning* sekitar 0,5, dan momentum 0,6.

Tabel 1. 1 Tabulasi Studi Pustaka dan Pengembangan yang dilakukan

No	Peneliti	Analisis Kasus	Metode	Target Analisis
1	(Nozaki & Nakamoto, 2017)	Melakukan analisis rekaman suara asli dan rekaman <i>voice changer</i> menggunakan teknik <i>autoencoder</i> atau <i>automatic statistic</i>	<i>Itakura-Saito Distance</i> berbasis <i>autoencoder</i>	Keakuratan asli atau identik dalam melakukan analisis suara
2	(Ishii dkk., 2013)	Menganalisis suara yang bergema dan berisik menggunakan teknik <i>autoencoder denoising</i>	<i>Autoencoder</i> berbasis <i>front-End</i>	Keakuratan pengenalan ucapan kosakata
3	(Enqvist & Karlsson, 2008)	Menganalisis rekaman suara pidato menggunakan <i>automatic statistic</i>	<i>Itakura-Saito Distance</i> dan <i>Apriori</i>	Mendapatkan perkiraan kedekatan <i>spectrum</i> rekaman suara
4	(Putri & Sunarno, 2014)	Menganalisis rekaman suara menggunakan teknik <i>audio forensic</i> untuk keperluan barang bukti digital	Analisis statistik <i>anova</i> dan <i>LR formant bandwidth</i>	Analisis <i>distribution</i> grafis dan analisis <i>spectrogram</i> pada enam rekaman suara tanpa manipulasi objek suara

Tabel 1. 2 Lanjutan

No	Peneliti	Analisis Kasus	Metode	Target Analisis
5	(Irawan, 2019)	Menganalisis keakuratan rekaman suara dengan membandingkan metode <i>Itakura-Saito Distance</i> dan metode konvensional (manual statistik)	Manual Statistik ( <i>Pitch, Formant, Spectrogram</i> ) dan <i>Itakura-Saito Distance</i>	Akurasi identifikasi suara pada <i>audio forensic</i> dengan perbandingan rekaman suara asli dan suara pembanding
6	(Hilmi, 2014)	Menerapkan kombinasi antara <i>digital signal processing</i> dengan <i>fast fourier transform</i> dan pengenalan pola suara dengan menggunakan metode jaringan syaraf tiruan ( <i>neural network</i> )	<i>Fast Fourier Transform</i> dan <i>Backpropagation Neural Network</i>	Mengidentifikasi suara laki-laki dan perempuan

### 1.7. Metode Penelitian

Penelitian ini perlu disusun langkah-langkah penyelesaian secara sistematis, adapun sistematisa metodologi yang digunakan pada penelitian ini adalah sebagai berikut:

#### 1. Tinjauan Pustaka

Tahap tinjauan pustaka dilakukan guna mencari bahan materi secara teoritis dan literatur dari penelitian-penelitian terdahulu, sehingga penelitian ini tidak terjadi perulangan pembahasan, dan menjadikan penelitian yang efektif ataupun bernilai.

#### 2. Teknik Pengambilan Sampel Data

Selanjutnya dari berbagai penelitian yang telah ada dirumuskan hal-hal yang penting pada *audio forensic*. Cara pengambilan sampel data suara sesuai dengan prosedur dan parameter yang ada, sehingga memperoleh pemahaman yang komprehensif mengenai metode yang diangkat.

#### 3. Implementasi Metode



Indikator dan parameter yang ditemukan dilakukan uji coba perumusan perbandingan dari masing-masing komponen yang telah ditentukan.

#### **4. Analisis Data**

Hasil dari implementasi metode tersebut akan dianalisis dan diolah sehingga menghasilkan nilai keakuratan dari implementasi metode tersebut.

#### **5. Laporan**

Semua kegiatan yang dilakukan dalam penelitian ini dibuat laporan dan didokumentasi dari kesimpulan analisis, baik kelebihan atau kekurangan metode yang digunakan. Sehingga bisa digunakan sebagai karya ilmiah untuk dijadikan bahan acuan dan referensi penelitian berikutnya.

### **1.8. Sistematika Penelitian**

Untuk memberikan gambaran dan mempermudah dalam penyusunan penelitian ini, maka dibuat sistematika penulisan sebagai berikut:

#### **BAB I PENDAHULUAN**

Pendahuluan merupakan pengantar terhadap permasalahan yang akan dibahas, didalamnya menguraikan tentang gambaran suatu penelitian yang terdiri dari latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian, dan sistematika penelitian.

#### **BAB II TINJAUAN PUSTAKA**

Pada bab ini menjelaskan tentang teori-teori dari berbagai bahan referensi yang terkait guna memecahkan masalah dalam penelitian ini.

#### **BAB III METODOLOGI PENELITIAN**

Pada bab ini membahas tentang langkah-langkah atau prosedur penelitian, sehingga mampu menyelesaikan masalah secara sistematis.

#### **BAB IV PEMBAHASAN**

Pada bab ini membahas tentang hasil dan pembahasan, terkait dengan pembahasan penyelesaian masalah yang diangkat, penentuan hasil analisis, dan evaluasi dari penelitian yang diangkat.

#### **BAB V PENUTUP**

Bab ini merupakan bab terakhir yang memuat kesimpulan akhir dari semua proses penelitian sampai kepada hasil implementasi metode dan saran yang perlu diperhatikan.

## BAB 2

### Tinjauan Pustaka

#### 2.1. Audio

*Audio* adalah suara yang dihasilkan oleh getaran benda atau suatu benda yang berupa sinyal analog dengan *amplitude* yang berubah secara terus menerus terhadap waktu dan suara yang berkaitan dengan indra pendengaran yang merambat melalui udara. Bila memperhatikan *gender*, bahwa *gender* perempuan memiliki tingkat keberhasilan jauh lebih tinggi dibandingkan dengan laki-laki kalau berkaitan dengan suara yang dibandingkan. Sehingga, apabila melakukan pengujian adalah suara perempuan, sistem akan mudah mengenalinya sebagai suara dari *user* yang tidak berhak melakukan akses ke sistem. Sebaliknya, sistem masih memiliki kesulitan dalam membedakan suara yang diuji, apabila suara tersebut merupakan suara dari *user* yang memiliki *gender* laki-laki (Imario, dkk., 2017).

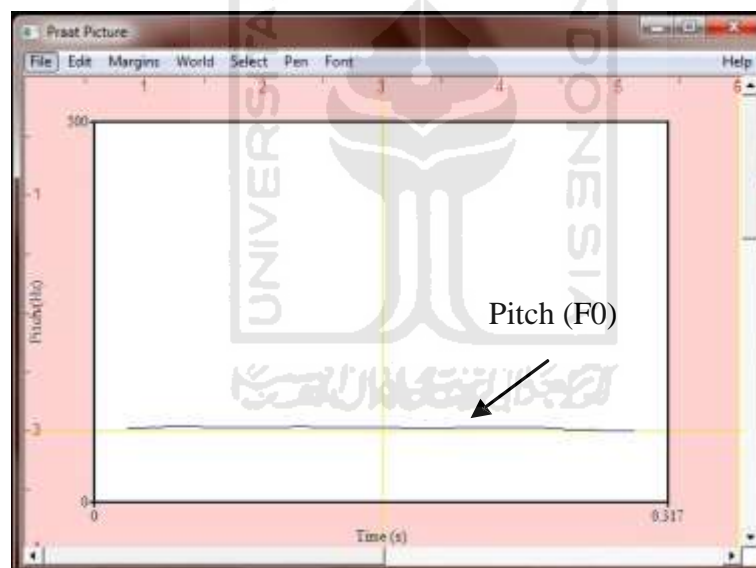
Suara dihasilkan melalui dua buah proses yaitu *Generation* dan *Filtering*. Pada proses *Generation*, suara pertama kali akan diproduksi melalui bergetarnya pita suara (*vocal cord* dan *vocal fold*) yang berada di *larynx* untuk menghasilkan bunyi periodik. Bunyi periodik yang sifatnya konstan tersebut kemudian disaring melalui *vocal tract* (juga disebut dengan istilah *resonator* suara atau *articulator*) yang terdiri dari lidah (*tongue*), gigi (*teeth*), bibir (*lips*), langit-langit (*palate*) dan lain-lain sehingga bunyi tersebut dapat menjadi bunyi keluaran (*output*) berupa bunyi vokal (*vowel*) dan atau bunyi konsonan (*consonant*) yang membentuk kata-kata yang memiliki arti yang nantinya dapat dianalisis untuk *voice recognition* (Al-Azhar Nuh, 2011). *Audio* yang berupa rekaman suara memiliki karakter pengucapan kosakata yang berbeda-beda dan kondisi yang tidak jelas. Rekaman suara berisi gelombang bunyi yang direkam dengan teknologi *digital recording* (Mansyur, 2017).

Menurut (Al-Azhar Nuh, 2011) rekaman suara dimungkinkan mengandung *noise*. Dalam pengertian umum, *noise* adalah suatu gangguan yang “didengar” orang lain, namun dalam istilah telekomunikasi kata *noise* juga dipakai untuk istilah gangguan yang menimbulkan kebisingan yang dapat didengar suatu sistem. *Noise* dapat terjadi dengan berbagai macam cara, misalnya rekaman suara barang bukti terdapat bocoran yang terjadi karena pada saat rekaman diperoleh tersangka berada pada lokasi dimana terdapat suara-suara yang saling bercampur. Pada prinsipnya, *noise* jenis ini tidak dapat dihilangkan tetapi dapat dikurangi keberadaan pada sumbernya barang bukti rekaman suara.

Menurut Al-Azhar Nuh (2011) pada prinsipnya suara terdiri dari banyak komponen yaitu *pitch*, *formant*, dan *spectrogram* yang dapat digunakan untuk mengidentifikasi karakteristik suara seseorang untuk kepentingan *voice recognition*. Berikut penjelasan dari masing-masing komponen yang disebutkan:

### 1. *Pitch*

Frekuensi getar dari pita suara yang disebut dengan istilah notasi  $F_0$ . Masing-masing orang memiliki frekuensi khas (*habitual pitch*) yang sangat dipengaruhi oleh berbagai aspek. Pada kondisi pembicaraan normal, level *habitual pitch* berkisar pada 50 s/d 250 Hz untuk laki-laki dan 120 s/d 500 Hz untuk perempuan. Frekuensi  $F_0$  ini berubah secara konstan dan memberikan informasi linguistik seseorang seperti pembeda antara intonasi dan emosi. Analisis *pitch* dapat digunakan untuk melakukan *voice recognition* terhadap suara seseorang yaitu melalui analisis statistik terhadap nilai *minimum pitch*, *maximum pitch* dan *mean pitch*. Diagram *pitch* dapat pada Gambar 2. 1.

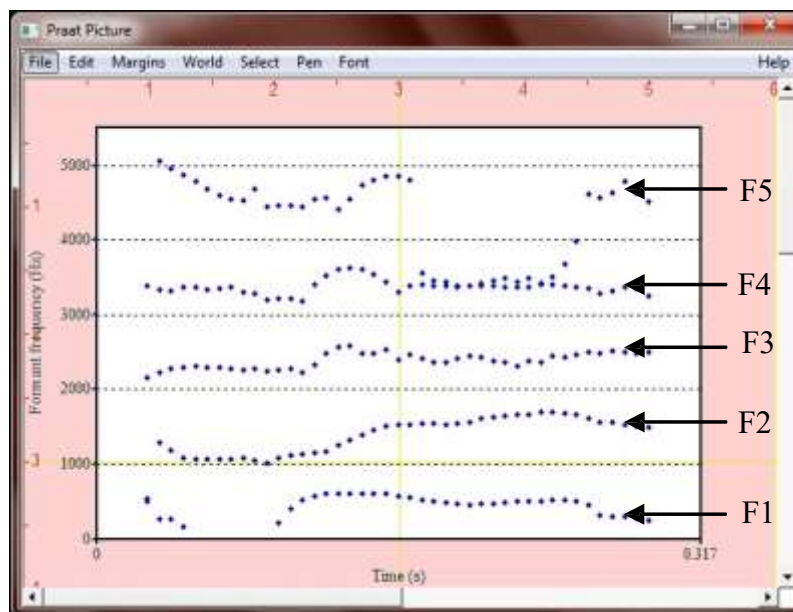


Gambar 2. 1 Diagram *pitch* terhadap waktu berubah secara konstan

(Sumber:(Al-Azhar Nuh, 2011))

### 2. *Formant*

*Formant* adalah frekuensi-frekuensi resonansi dari filter, yaitu *vocal tract* yang meneruskan dan memfilter bunyi berupa kata-kata. Secara umum, frekuensi-frekuensi *formant* bersifat tidak terbatas, namun untuk mengidentifikasi seseorang paling tidak ada 3 (tiga) format yang dianalisis yaitu, *Formant 1* ( $F_1$ ), *Formant 2* dan *Formant 3* ( $F_3$ ). Diagram *formant* dapat dilihat pada Gambar 2. 2.

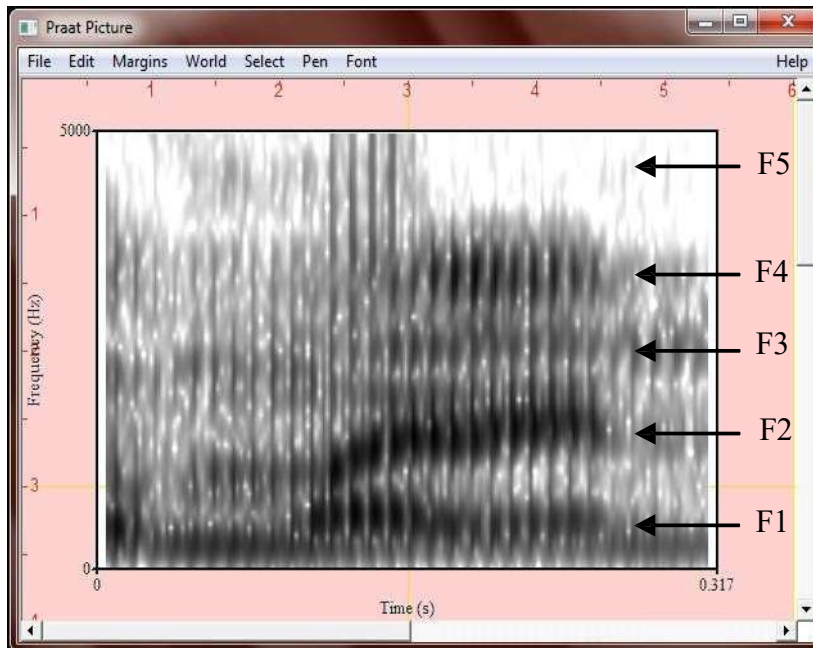


Gambar 2. 2 Diagram masing-masing Formant F1, F2, F3, F4 dan F5.

(Sumber:(Al-Azhar Nuh, 2011))

### 3. *Spectrogram*

*Spectrogram* merupakan representasi *spectral* yang bervariasi terhadap waktu yang menunjukkan tingkat intensitas energi *spectral*, membentuk pola umum yang khas dalam pengucapan kata, dan pola khusus masing-masing *formant* dalam pengucapan suku kata. Dengan kata lain *spectrogram* adalah bentuk visualisasi dari masing-masing nilai *formant* yang dilengkapi dengan level energi yang bervariasi terhadap waktu. Level energi ini dikenal dengan istilah *formant bandwidth*. Nantinya pada kasus-kasus yang bersifat pemalsuan suara dengan teknik *pitch shift* atau pelaku berusaha untuk menghilangkan karakter suara aslinya, maka *formant bandwidth* dapat digunakan untuk memetakan atau mengidentifikasi suara aslinya. Jika durasi rekaman suara *unknown* lumayan panjang, maka analisis *spectrogram* juga dapat digunakan untuk mempercepat pemilihan pengucapan kata-kata yang akan dianalisis dalam rangka untuk mendapatkan jumlah minimal 20 kata untuk dapat menunjukkan keidentikan suara *unknown* dengan *known* atau pembanding. Representasi *spectrogram* dapat dilihat pada Gambar 2. 3.



Gambar 2. 3 Representasi spectrogram spektral dengan tingkatan energinya.  
(Sumber:(Al-Azhar Nuh, 2011))

## 2.2. Audio Forensic

*Audio forensics* merupakan penerapan ilmu pengetahuan dan metode secara ilmiah dalam penanganan barang bukti yang berupa *audio*, dimana barang bukti ini rentan dirubah dan dimanipulasi (Al-Azhar Nuh, 2011). Keaslian barang bukti *audio* dalam menguatkan keputusan hakim dipersidangan dilakukan dengan serangkaian pengamatan dan tes untuk mengevaluasi integritas rekaman (Mansyur, 2017).

Sebagaimana yang disampaikan Al-Azhar Nuh (2011) Mengidentifikasi rekaman suara diperlukan verifikasi dengan suara pembanding yang hampir mirip, sehingga perlu adanya tahapan-tahapan sesuai dengan SOP tentang analisa *audio forensics* pada DFAT PUSLABFOR yang mengacu pada *Good Practice Guide for Computer-Based Electronic Evidence* yang diterbitkan oleh ACPO di Inggris, dan *Forensic Examination of Digital Evidence: A Guide for Law Enforcement* yang diterbitkan oleh *National Institute of Justice* yang berada di bawah *Department of Justice*, Amerika Serikat. Analisa yang harus dilakukan pada *audio forensics* sebagai berikut:

### 1. Acquisition

Proses dimana pengambilan barang bukti asli yang selanjutnya dicatat menggunakan teknik *audio recorder* untuk mendapatkan model, ukuran, dan *serial number*. Rekaman suara asli seharusnya di *backup* terlebih dahulu, supaya tetap menjaga nilai keaslian dari

barang bukti yang ditemukan. Salah satu tahap pemeriksaan adalah mendapatkan suara pembanding terhadap suara asli yang ada di *audio recorder* dengan dilengkapi administrasi penyidikan yang lengkap. Proses akuisisi *audio recorder* akan menghasilkan *file DD image*.

## 2. Peningkatan *Audio (audio enhancement)*

Hasil *backup* dari barang bukti rekaman suara asli diputar berkali-kali untuk melihat kualitas rekaman, jika suara yang didengar tidak bagus dikarenakan banyak *noise*. Suara yang demikian perlu adanya peningkatan untuk menaikkan kualitas rekaman, sehingga kosakata pembicaraan bisa jelas didengar.

## 3. *Decoding*

Setelah rekaman suara sudah bisa didengar kosakata dengan jelas, selanjutnya dilakukan pemilihan antara kosakata yang jelas dan tidak. Kemudian suara dengan kosakata yang jelas dicatat untuk dijadikan transkrip rekaman.

## 4. *Voice Recognition*

Proses menganalisis suara yang ada adalah identik atau tidak dengan mengambil kosakata yang jelas dari transkrip rekaman yang asli, selanjutnya melakukan perbandingan suara lain yang hamper mirip dengan kosakata yang sama. Analisis menggunakan parameter terhadap *pitch*, *formant bandwidth*, dan *spectrogram*. Kosakata yang didapat minimal dua puluh (20) kata yang memiliki kesamaan antara rekaman suara asli dan pembanding, jika kurang dari dua puluh kata maka tidak memenuhi syarat *audio forensic*.

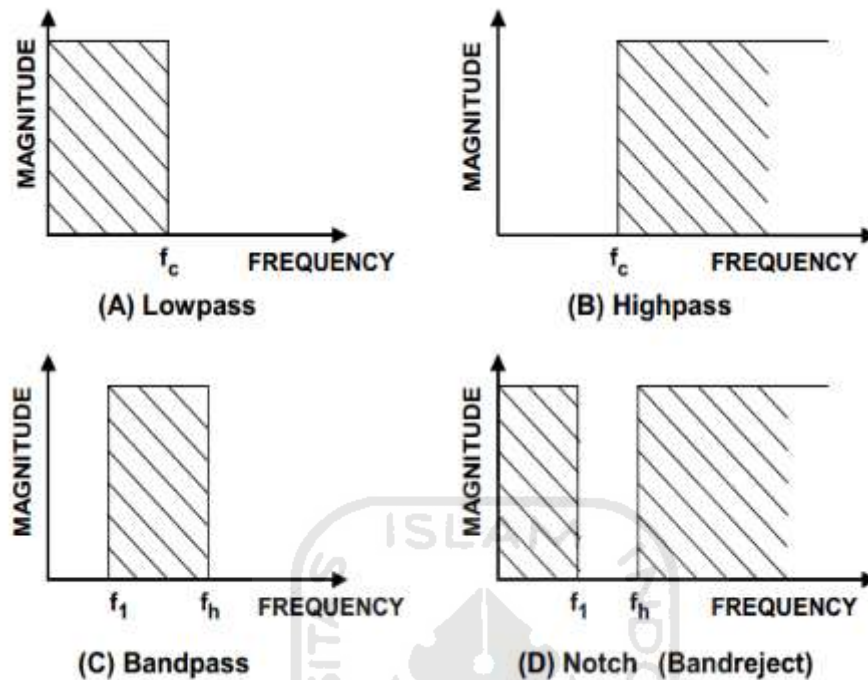
### 2.3. *Lowpass Filter*

Menurut Singh & Kumar Garg (2014) *filtering* secara efektif dapat mengurangi komponen frekuensi yang tidak diinginkan dalam sebuah sinyal atau dikenal istilah *noise*. Adapun jenis-jenis *noise* yaitu:

1. Elektronik *noise* seperti *thermal noise* dan *shot noise*
2. Suara akustik yang berasal dari sumber yang bergerak, bergetar, dan bertabrakan
3. *Electromagnetic noise*
4. *Electrostatic noise* yang berasal dari tegangan
5. Saluran komunikasi distorsi dan *fading*
6. *Lost control network* yang bisa menyebabkan paket data hilang

Dalam lingkungan yang sangat bising menghilangkan *noise* dapat memulihkan sinyal suara asli, sehingga kosakata akan terdengar lebih jelas. *Filter* memiliki banyak jenis yaitu

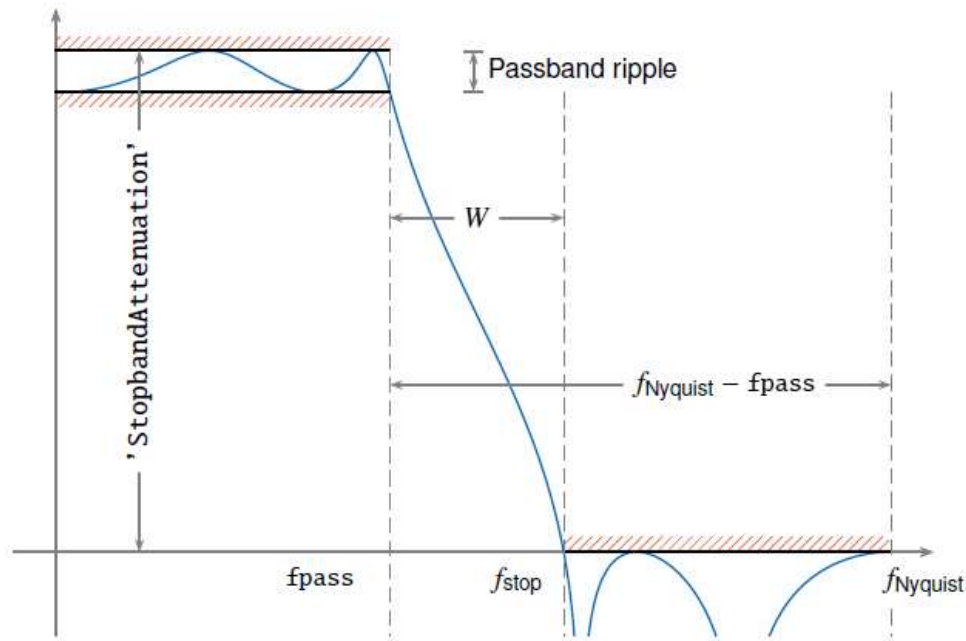
*lowpass*, *highpass*, *bandpass*, dan *notch*. *Lowpass filter* seringkali digunakan untuk menstabilkan *amplifier* dengan menghilangkan *high frequency* yang masih terdapat pada *noise*. Berikut jenis-jenis dari *filter* dapat dilihat pada Gambar 2. 4.



Gambar 2. 4 Jenis-jenis *filter*

(Sumber: ( Singh & Kumar Garg (2014))

Sampel suara yang diambil dengan format *.wav* dimungkinkan melakukan pengurangan *noise* untuk mendapatkan frekuensi sinyal yang efisien (Salih, 2017). Untuk menafsirkan *lowpass filter* dapat dilihat pada Gambar 2. 5.



Gambar 2. 5 low-pass filter specifications

(Sumber: Mathworks.com)

*Nyquist frequency*,  $f_{Nyquist}$  adalah komponen frekuensi tinggi dari sinyal yang mendapatkan sampel pada tingkat tertentu.  $f_{Nyquist} = 1(\frac{x\pi rad}{sample})$  ketika sinyal input tidak memiliki informasi waktu,  $\frac{f_s}{2}$  hertz ketika sinyal input adalah laju sampel. *Stopband frequency*,  $f_{stop}$  adalah frekuensi diluar redaman yang sama atau lebih besar dari nilai yang ditentukan. *Transition width* adalah  $W$ , is  $f_{stop} - f_{pass}$ . Saat menentukan nilai  $s$  untuk *steepness* fungsi dapat dilihat pada persamaan (1).

$$W = (1 - d)x(f_{Nyquist} - f_{pass}) \quad (1)$$

#### 2.4. Fast Faurier Transform (FFT)

FFT dapat digunakan untuk menganalisis sinyal domain frekuensi, dimana karakter pergerakan berada di *low spectrum*. Parameter pergerakan berbasis FFT termasuk informasi intensitas yang memiliki korelasi yang lebih tinggi satu sama lain (Timotius et al., 2018). Frekuensi merupakan jumlah getaran suara yang dalam satu detik. Pergerakan sinyal berasal *amplitude* pada sebuah *file* suara dalam interval waktu tertentu yang sering disebut sinyal analog (Prasetya, dkk., 2011). Metode FFT merupakan pengembangan dari metode DFT (*Discrete Fourier Transform*). DFT merupakan metode untuk mengubah sinyal analog yang berbasis waktu menjadi berbasis *frequency* (Alfina, dkk, 2019).



Sebagaimana disampaikan Batenkov (2005) DFT untuk *finite-length sequence*  $y(n)$  yang terdefinisi rentang  $0 \leq n \leq N - 1$  dinyatakan dalam persamaan (2):

$$y(t) = y(0)\delta(0) + y(T_s)\delta(t - T_s) + y(2T_s)\delta(t - 2T_s) + \dots + y((N - 1)T_s)\delta(t - (N - 1)T_s) \quad (2)$$

$$\text{Dimana } \frac{1}{T_s} \geq 2f_{max}$$

## 2.5. Akurasi Suara

Perolehan hasil akurasi tidak lepas dari analisa anova. Analisa *anova* adalah analisa yang mengkalkulasi secara statistik nilai-nilai *formant* dari suara *unknown* dan suara *known*. Kalkulasi nilai *formant* menggunakan *Likelihood Ratio* yang dapat dilihat pada persamaan (3).

$$LR = \frac{p(E|H_p)}{p(E|H_d)} \quad (3)$$

Dimana  $p(E|H_p)$  berasal dari *p - value anova* merupakan hipotesis tuntutan (*prosecution*), yaitu sampel suara *known* dan *unknown* berasal dari orang sama. Sedangkan  $p(E|H_d) = 1 - p(E|H_p)$  merupakan hipotesis perlawanan, yaitu suara *known* dan *unknown* berasal dari orang yang berbeda. Jika  $LR > 1$  maka mendukung  $p(E|H_p)$ , sebaliknya jika  $LR < 1$  maka  $p(E|H_d)$  yang didukung. Untuk itu nilai  $p(E|H_p) > 0.5$  untuk dapat menyimpulkan suara *known* dan *unknown* identik (Al-Azhar, 2012).

## 2.6. Itakura Saito Distance

Menurut Gray (2007) Metode *Itakura Saito* diusulkan oleh Fumitada Itakura dan Shuzo Saito pada 1960 untuk mengukur kemiripan suara. Metode ini cukup memegang peranan dalam hal kajian mengenai *speech coding, analysis, synthesis and recognition*. *Itakura Saito* dikenal juga dengan *maximum likelihood distortion measure*, dimana *linear prediction* (LP) dijadikan perkiraan kemungkinan. *Linear prediction* yaitu bagaimana mengoptimalkan kecepatan *bit* yang rendah dengan kualitas suara memadai.

*Optimal linear prediction* pada *Itakura Saito autocorrelation* dengan analisis proses *Gaussian*, dimana cara memperkirakan korelasi data dari sinyal suara secara otomatis diamati dengan model diskrit. Misalnya parameter  $a = (a_0 = 1, a_1, \dots, a_m)$  dapat dihitung dengan *random vector* ( $X_m$ ). Prinsipnya asumsi *Gaussian* dalam mencari *invers* parameter  $a$  menggunakan *Cholesky decomposition*. Sedangkan untuk mencari  $R_{m+1}$  menggunakan

matriks *Toeplitz* dengan urutan dimensinya  $(a + 1)(a + 1)$ , sehingga dapat dilihat pada persamaan (4).

$$\begin{aligned} \sim X_m &= -\sum_{l=1}^m a_l X_{m-l} \text{ dimana} \\ \text{MSE } E[\epsilon_m^2] &= a^t R_{m+1} a \end{aligned} \quad (4)$$

Sebagaimana disampaikan Padhy & Sahu (2009) model dari *Linear Prediction Coding* (LPC) adalah sampel dari suara asli yang diberikan pada waktu  $n$ ,  $s(n)$ , mendekati jarak ( $p$ ) suara pembanding, sehingga dapat dilihat pada persamaan (5).

$$s(n) = a_1 s(n - 1) + a_2 s(n - 2) + \dots + a_p s(n - p) \quad (5)$$

Dimana koefisien dari parameter ( $a$ ) konstan terhadap panjang *frame* dari suara yang dianalisis, sehingga mengubah persamaannya dan menambah *variable*  $G u(n)$  seperti persamaan (6).

$$s(n) = \sum_{i=1}^p a_i s(n - i) + G u(n) \quad (6)$$

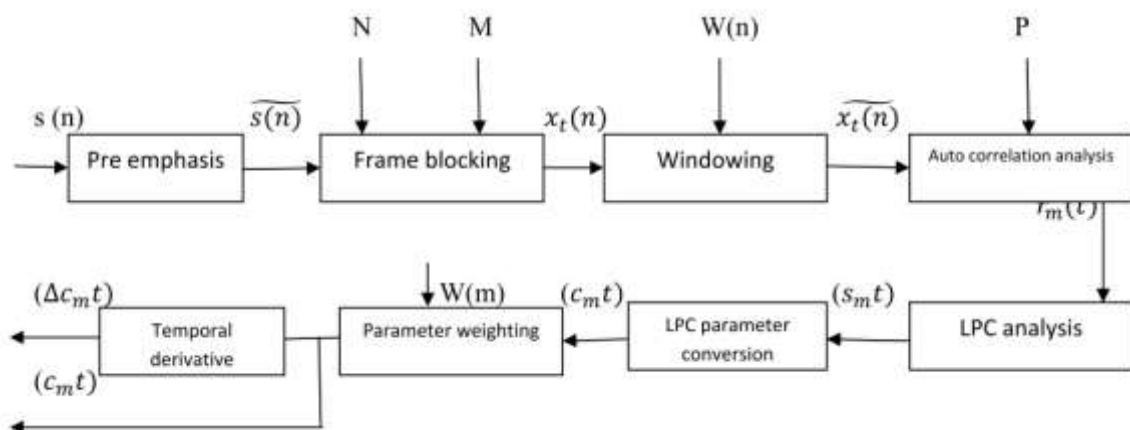
Dimana  $G u(n)$  adalah normalisasi *output* dari waktu ( $n$ ) dan menambahkan domain  $z$  untuk dijadikan relasi, sehingga dapat dilihat pada persamaan (7).

$$S(z) = \sum_{i=1}^p a_i z^{-i} S(z) + G u(z) \quad (7)$$

Sehingga dari persamaan diatas dapat ditarik kesimpulan pada persamaan (8).

$$H(z) = \frac{S(z)}{G u(z)} = \frac{1}{1 - \sum_{i=1}^p a_i z^{-i}} = \frac{1}{A(z)} \quad (8)$$

Persamaan 14 menunjukkan sumber yang dinormalisasi pada  $u(n)$  dan skala di *gain*  $G$ , sedangkan yang bertindak sebagai *input* adalah  $H(z) = \frac{1}{A(z)}$  untuk menghasilkan sinyal suara  $s(n)$ . Berikut adalah diagram dari analisis LPC dapat dilihat pada Gambar 2. 6.



Gambar 2. 6 Diagram blok LPC untuk *speech recognition*

Gambar 2. 6 menjelaskan alur proses dari LPC untuk mengekstraksi *frame* dari sinyal suara dari parameter yang terpisah. Terdapat tiga hal utama dari proses LPC yaitu pertama *pre emphasis* dimana memproses sinyal suara pertama pada domain waktu dan meratakan sinyal secara *spectral*, dapat dilihat pada persamaan (9).

$$\sim A(n) = A(n) - a * A(n - 1) \quad (9)$$

*Frame blocking* memberikan jumlah sampel ucapan pada  $N$  (misal 300) berturut-turut dan *frame* diberikan jarak pada  $M$  (misal 100) dengan sampel suara terpisah. ( $\leq L - 1$ ) menyiratkan jaraknya tidak negatif dan meminimalkan *Itakura Saito* yang menghasilkan nilai 0, sehingga dapat dilihat pada persamaan (10).

$$X, (n) = A(M * 1 + n, 0 \leq n \leq N - 1 \text{ and } 0 \leq 1 \leq L - 1) \quad (10)$$

*Frame windowing* dimana setiap sampel  $N$  dikalikan sampel *hamming window*  $W(n)$ . *Hamming window* digunakan untuk meminimalkan efek buruk saat mengurangi *noise* dari sinyal bagian sampel  $N$ , sehingga dapat dilihat pada persamaan (11).

$$\sim U, (n) = X, (n) * W(n), 0 \leq n \leq N - 1 \quad (11)$$

Perbedaan *log spectral*  $V(w)$  adalah dasar dari banyak tindakan distorsi ucapan, sehingga *Itakura Saito* menjadi formulasi *linear prediction* dengan perkiraan-perkiraan yang ada.  $d_{1S}$  adalah variabel dari *Itakura Saito* yang dapat dilihat pada persamaan (12).

$$\begin{aligned} d_{1S}(S, S') &= \int_{-\pi}^{\pi} (e^{V(w)} - V(w) - 1) \frac{dw}{2\pi} \\ &= \int_{-\pi}^{\pi} \frac{S(w)}{S'(w)} \frac{dw}{2\pi} - \log \frac{\sigma_{\infty}^2}{\sigma_{\infty}^{\prime 2}} - 1 \end{aligned} \quad (12)$$

Dimana  $\sigma_{\infty}^2$  *spectrum* sinyal rekaman suara *unknown* dan  $\sigma_{\infty}^{\prime 2}$  *spectrum* sinyal rekaman suara *known* yang merupakan *prediction errors* pada  $S(w) S'(w)$ . *Maximum likelihood* dari *Itakura Saito* merupakan langkah-langkah distorsi ucapan pada *spectrum* suara asli dan perkiraan.

## 2.7. Jaringan Syaraf Tiruan

Dikutip dari geeksforgeeks.org studi tentang jaringan syaraf tiruan tentu terinspirasi oleh pengamatan bahwa pembelajaran biologis dibangun dari jaringan yang sangat kompleks dari *neuron* yang saling berhubungan dengan otak. Otak manusia mengandung jaringan yang sekitar  $10^{11} - 10^{12}$  *neuron* yang rata-rata terhubung ke  $10^1$  *neuron* untuk membuat keputusan yang sangat kompleks. Model awal *neuron* buatan diperkenalkan oleh Warren McCulloch dan Walter Pitts pada 1943 yang dikenal dengan *linear threshold gate*. Hal itu

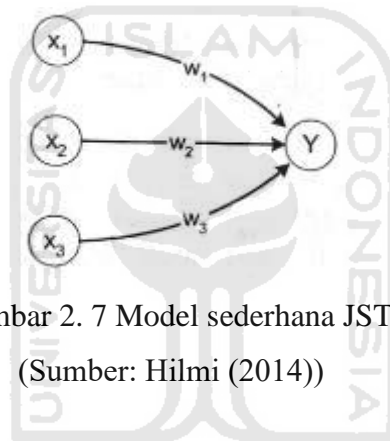
dikenal dengan *neuron* yang mempunyai serangkaian *input*  $I_1, I_2, \dots, I_i$  dan satu *output*  $y$  (biner). Secara matematis dapat dilihat pada persamaan (13).

$$\begin{aligned} \text{Sum} &= \sum_{i=1}^N I_i W_i \\ y &= f(\text{Sum}) \end{aligned} \quad (13)$$

$W_1, W_2, \dots, W_i$  adalah nilai bobot yang dinormalisasi dalam kisaran  $(0,1)$  atau  $(-1,1)$  dan terkait dari setiap *input* ( $I$ ). *Sum* adalah jumlah bobot dan merupakan *threshold constant*. Fungsi  $f$  adalah fungsi langkah *linear* di *threshold*.

JST ditentukan oleh tiga hal yaitu:

1. Pola hubungan antar *neuron*
2. Metode untuk menentukan bobot penghubung (metode *training* atau *learning*)
3. Fungsi aktivasi, berikut model sederhana dari jaringan syaraf tiruan dapat dilihat pada Gambar 2. 7.



Gambar 2. 7 Model sederhana JST  
(Sumber: Hilmi (2014))

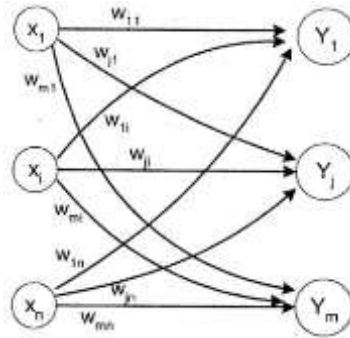
Sebagai contoh *neuron*  $y$  menerima *input* dari *neuron*  $x_1, x_2, x_3$  dengan bobot saling berhubungan  $w_1, w_2, w_3$ , kemudian menjadi persamaan (14).

$$\text{net} = x_1 w_1 + x_2 w_2 + x_3 w_3 \quad (14)$$

Besarnya *impuls* yang diterima oleh  $y$  mengikuti fungsi aktivasi  $y = f(\text{net})$ . Apabila nilai fungsi aktivasi cukup kuat, maka sinyal diteruskan. Nilai fungsi aktivasi juga dapat dipakai sebagai dasar merubah bobot. JST mempunyai arsitektur jaringan yaitu:

1. *Single layer network*

Jaringan ini terdapat sekumpulan *input neuron* yang dihubungkan langsung dengan sekumpulan *output*. Dalam beberapa kasus, hanya ada sebuah unit *neuron output*. Besarnya bobot terhubung antara unit  $ke - i$  dalam unit  $ke - j$  dalam *output*, sehingga selama proses pelatihan, bobot-bobot tersebut akan dimodifikasi untuk meningkatkan akurasi. Arsitektur *single layer network* dapat dilihat pada Gambar 2. 8.

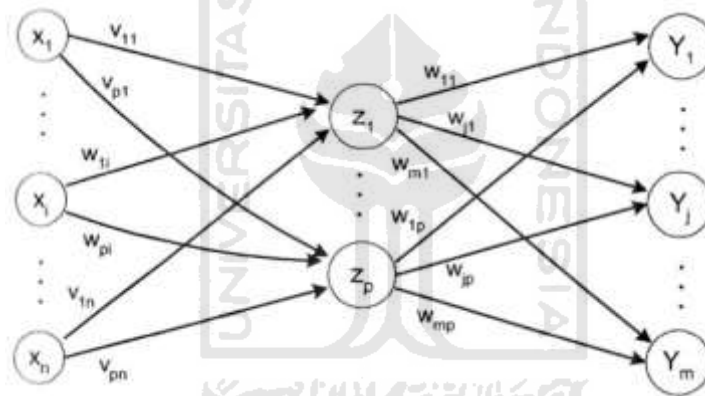


Gambar 2. 8 Arsitektur *single layer network*

(Sumber: Hilmi (2014))

## 2. *Multilayer network*

Jaringan ini merupakan perluasan dari *single layer*, dimana selain ada *input output* juga ada unit lain yang disebut *hidden layer*. *Multilayer network* dapat menyelesaikan masalah lebih kompleks dibandingkan dengan jaringan sebelumnya, meskipun prosesnya sedikit lama. Arsitektur *multiplayer network* dapat dilihat pada Gambar 2. 9.



Gambar 2. 9 Arsitektur *multiplayer network*

(Sumber: Hilmi (2014))

## 3. Model JST dua dengan *feedback*

Dengan jumlah *neuron* yang sangat besar, JST memiliki sifat yaitu *fault tolerance*. Sifat ini mengandung maksud kerusakan dari sebagian sel-sel dalam jaringan, tetapi tidak mempengaruhi *output* yang akan dikeluarkan. Model ini mempunyai sifat umpan balik, sehingga *output* yang dihasilkan akan mempengaruhi *input* dimasukkan lagi.

## 4. Model JST lapisan kompetitif

Bentuk dari jaringan ini merupakan bagian dari jumlah yang besar pada jaringan syaraf. Pada dasarnya, hubungan antar *neuron* pada lapisan kompetitif tidak ditunjukkan secara arsitektural pada jaringan syaraf.

Ada beberapa fungsi aktivasi yang sering digunakan dalam syaraf tiruan. Fungsi aktivasi yang digunakan pada *backpropagation* antara lain:

1. Fungsi *sigmoid biner*

Fungsi *sigmoid biner* memiliki nilai pada *range* 0 sampai 1, oleh karena itu fungsi ini sering digunakan untuk jaringan syaraf yang membutuhkan nilai *output* yang terletak pada interval 0 sampai 1. Fungsi *sigmoid biner* dapat dilihat pada persamaan (15).

$$y = f(x) = \frac{1}{1+e^{-\sigma x}} \tag{15}$$

Dengan:

$$f'(x) = \sigma f(x)(1 - f(x))$$

2. Fungsi *sigmoid bipolar*

Fungsi ini hampir sama dengan *sigmoid biner*, hanya saja *output* dari fungsi ini memiliki *range* antara 1 sampai -1. Fungsi *sigmoid bipolar* dapat dilihat pada persamaan (16).

$$y = f(x) = \frac{1-e^{-x}}{1+e^{-x}} \tag{16}$$

Dengan:

$$f'(x) = \frac{\sigma}{2} [1 + f(x)][1 - f(x)]$$

3. Fungsi *linear*

Fungsi *linear* memiliki nilai *output* yang sama dengan nilai *input*, sehingga dapat dilihat pada persamaan (17).

$$y = x \tag{17}$$

Sebagaimana yang disampaikan Bre, Gimenez, & Fachinotti ( 2018) JST adalah sekumpulan satu unit *neuron* yang tersusun dalam *input layer*, *hidden layer*, dan *output layer*. Setiap *neuron* terhubung ke *neuron-neuron* di *neighboring layer* melalui bobot adaptif. Pada persamaan (18) menunjukkan model *neuron j* genetik pada *hidden layer k*, yang hasilnya didefinisikan sebagai berikut:

$$y_j^k = f \sum_{i=1}^n w_{ij}^k x_i^k + b_j^k \tag{18}$$

## 2.8. Backpropagation

Menurut Sidqi (2018) *Backpropagation* adalah algoritma untuk memperkecil tingkat *error* dengan cara menyesuaikan bobot melalui perbedaan *output* dan target yang diinginkan. *Backpropagation* sering dikatakan algoritma *multilayer*, yang mempunyai parameter yaitu *input layer*, *hidden layer*, dan *output layer*. Pada *input layer* tidak terjadi proses perhitungan,

tetapi terjadi pengiriman sinyal *input* ke  $x$  *hidden* layer. *Hidden* layer pada *backpropagation* dapat menyebabkan tingkat *error* menjadi lebih kecil. Hal itu disebabkan sebagai tempat untuk memperbaharui dan menyesuaikan bobot yang diinginkan. Bias diarahkan mendekati dengan target *output* yang diinginkan.

Algoritma *Backpropagation* adalah *gradien* dari fungsi biaya dihitung menurut bobot dan dalam bias jaringan untuk meminimalkan fungsi biaya. Efek perbaikan adalah hasil menggunakan rasio dari nilai target untuk menghitung gradien dari dua perbandingan nilai. Dengan kata lain, gradien yang berasal dari nilai *error* menghilang ketika kedua target dan *output* menjadi relatif kecil, sehingga perlambatan tidak bisa untuk *input* sasaran yang terkecil (Nozaki & Nakamoto, 2017).

Algoritma *backpropagation* menggunakan fungsi *sigmoid biner*, karena *output* yang diharapkan bernilai 0 sampai 1. Sebagaimana yang disampaikan (Sutojo, dkk, 2011) secara umum algoritma ini terdiri dari 3 langkah utama yaitu:

1. Tahap perambatan maju

Sebelum memulai tahap ini hal yang paling yaitu melakukan inisialisasi bobot (bobot diberi nilai acak yang bernilai kecil). Pola masukan dihitung maju mulai dari *input* layer hingga *output* layer menggunakan fungsi yang ditentukan.

- a. Setiap bagian *input* ( $X_i, i = 1, 2, 3, \dots, n$ ) menerima sinyal  $X_i$  dan meneruskan sinyal tersebut kesemua bagian pada lapisan tersembunyi.
- b. Setiap bagian tersembunyi ( $Z_j, j = 1, 2, 3, \dots, p$ ) menjumlahkan bobot sinyal *input* dengan persamaan (19).

$$Z_{in} = V_{0j} + \sum_{i=0}^n X_i V_{ij} \quad (19)$$

Penerapan fungsi aktivasi untuk menghitung sinyal *output* dengan persamaan (20).

$$Z_j = f(Z_{in\_j}) \quad (20)$$

Setiap bagian *output* ( $Y_k, k = 1, 2, 3, \dots, m$ ) menjumlahkan bobot sinyal *output* dengan persamaan (21).

$$Y_{in\_k} = W_{0k} + \sum_{j=0}^p Z_j W_{jk} \quad (21)$$

Dan menerapkan fungsi aktivasi untuk menghilang sinyal *output* dengan persamaan (22).

$$Y_k = f(Y_{ink}) \quad (22)$$

2. Tahap perambatan balik

Selisih antara keluaran jaringan dengan target yang diinginkan merupakan kesalahan yang terjadi. Kesalahan yang terjadi itu dipropagasi mundur, dimulai dari garis yang berhubungan langsung dengan bagian-bagian di *output* layer.

- a. Setiap bagian *output* ( $Y_k$ ,  $k = 1, \dots, m$ ) menerima pola target yang sesuai dengan pola *input* pelatihan, kemudian hitung *error* dengan persamaan (23).

$$\delta_k = (t_k - y_k)f'(y_{ink}) \quad (23)$$

$f'$  adalah turunan dari fungsi aktivasi, sedangkan parameter koreksi bias (dipakai untuk mengubah  $w_{ok}$ ), dihitung persamaan (24).

$$\begin{aligned} \Delta w_{jk}(t) &= \beta \delta_k z_j \\ \Delta w_{jk}(t+1) &= \beta \delta_k + \alpha \Delta w_{jk}(t) \end{aligned} \quad (24)$$

Sedangkan parameter koreksi bias (dipakai untuk mengubah  $w_{ok}$ ), dihitung dengan persamaan (25).

$$\begin{aligned} \Delta w_{ok}(t) &= \beta \delta_k \\ \Delta w_{ok}(t+1) &= \beta \delta_k + \alpha \Delta w_{ok}(t) \end{aligned} \quad (25)$$

Parameter diatas sekaligus mengirimkan  $\delta_k$  ke bagian-bagian yang ada dilapisan bagian kanan.

- b. Pada tiap unit tersembunyi ( $z_j$ ,  $j = 1, \dots, p$ ) menjumlahkan *delta input* ( $\delta_{in\_j}$ ) (dari bagian-bagian yang berbeda pada lapisan dikanannya) dihitung dengan persamaan (26).

$$\delta_{in\_j} = \sum_{k=0}^m \delta_k w_{jk} \quad (26)$$

Parameter untuk menghitung informasi *error*, kalikan nilai dengan turunan dari fungsi aktivitasnya dengan persamaan (27).

$$\delta_j = \delta_{in\_j} f'(z_{in\_j}) \quad (27)$$

Kemudian hitung koreksi bobot yang dipakai untuk mengubah  $v_{ij}$  dihitung dengan persamaan (28).

$$\begin{aligned} \Delta v_{ij}(t) &= \beta \delta_j x_i \\ \Delta v_{ij}(t+1) &= \beta \delta_j x_i + \alpha \Delta v_{ij}(t) \end{aligned} \quad (28)$$

Parameter koreksi bias yang dipakai untuk mengubah  $v_{oj}$  dihitung dengan persamaan (29).

$$\begin{aligned} \Delta v_{oj}(t) &= \beta \delta_j \\ \Delta v_{oj}(t+1) &= \beta \delta_j + \alpha \Delta v_{oj}(t) \end{aligned} \quad (29)$$

3. Tahap perubahan bobot dan bias



Modifikasi bobot untuk menurunkan kesalahan yang terjadi, ketiga fase tersebut diulang-ulang terus hingga kondisi penghentian dipenuhi.

- a. Setiap unit keluaran ( $y_k, k = 1, 2, \dots, m$ ) dilakukan perubahan bobot dan bias berdasarkan persamaan (30) dan persamaan (31) untuk ( $k = 0, 1, \dots, p$ ).

$$w_{jk}(t + 1) = w_{jk}(t) + \Delta w_{jk}(t + 1) \quad (30)$$

$$w_{0k}(t + 1) = w_{0k}(t) + \Delta w_{0k}(t + 1) \quad (31)$$

- b. Setiap bagian unit tersembunyi ( $z_j, j = 1, 2, \dots, p$ ) bias dan bobotnya diperbaharui dengan persamaan (32) dan persamaan (33) untuk ( $i = 0, 1, \dots, n$ ).

$$v_{ij}(t + 1) = v_{ij}(t) + \Delta v_{ij}(t + 1) \quad (32)$$

$$v_{oj}(t + 1) = v_{oj}(t) + \Delta v_{oj}(t + 1) \quad (33)$$

- c. Uji kondisi berhenti

Lamanya interaksi yang harus dilakukan menjadi masalah utama dalam algoritma *Backpropagation*. *Backpropagation* tidak dapat memberikan kepastian berapa nilai *epoch* yang harus dilalui untuk mencapai kondisi diinginkan, sehingga parameter yang ada dari setiap peneliti berbeda-beda dalam menghasilkan jumlah iterasi yang relatif lebih sedikit. Berikut penjelasan dari masalah yang pada *Backpropagation*.

1. Inisialisasi jaringan

Perintah yang dipakai dalam membentuk jaringan dapat dilihat pada persamaan 34.

$$net = newff(PR, [S1 \dots Sn], \{Tf1 \dots Tfn\}, Btf, Blf, Pf) \quad (34)$$

Dimana *net* merupakan jaringan *n layer*. *PR* merupakan matriks ordo  $R \times 2$ . *Sn* merupakan jumlah unit *n layer* (*tansig*). *Tfn* merupakan fungsi aktivasi yang dipakai *n layer* (*traingdx*). *Btf* merupakan fungsi pelatihan jaringan (*learnqdm*). *Blf* merupakan fungsi perubahan bobot/bias. *Pf* merupakan fungsi perhitungan *error* (MSE)

2. Inisialisasi bobot awal secara *random*

Pemilihan bobot awal sangat mempengaruhi terhadap nilai *error*, serta cepat tidaknya proses pelatihan. Apabila nilai bobot awal terlalu besar, maka *input* ke setiap *hidden layer* dan *output layer* akan jatuh pada daerah fungsi *sigmoid* yang sangat kecil. Sebaliknya, apabila nilai bobot awal terlalu kecil, maka *hidden layer* dan *output layer* akan sangat kecil, sehingga menyebabkan proses pelatihan berjalan lambat. Bobot awal diinisialisasi secara *random* dengan nilai antara -0.5 sampai 0.5 interval lainnya.

3. Jumlah unit tersembunyi (*hidden layer*)

Penentuan jumlah *hidden layer* kadangkala membuat pelatihan *Backpropagation* menjadi lebih mudah. Dalam perambatan maju, keluaran harus dihitung untuk tiap *layer*, dimulai dari *hidden layer* paling bawah (terdekat dengan *input*). Dalam perampatan mundur, faktor  $\delta$  perlu dihitung untuk tiap *hidden layer*, dimulai dari *output layer*.

#### 4. Jumlah pola pelatihan

Tidak ada kepastian tentang berapa banyak pola yang diperlukan dalam pelatihan. Jumlah pola yang dibutuhkan mempengaruhi banyaknya bobot dalam jaringan serta tingkat akurasi yang diharapkan. Aturan sederhana bisa ditentukan dengan persamaan berikut:

$$\text{jumlah pola} = \frac{\text{jumlah bobot}}{\text{tingkat akurasi}}$$

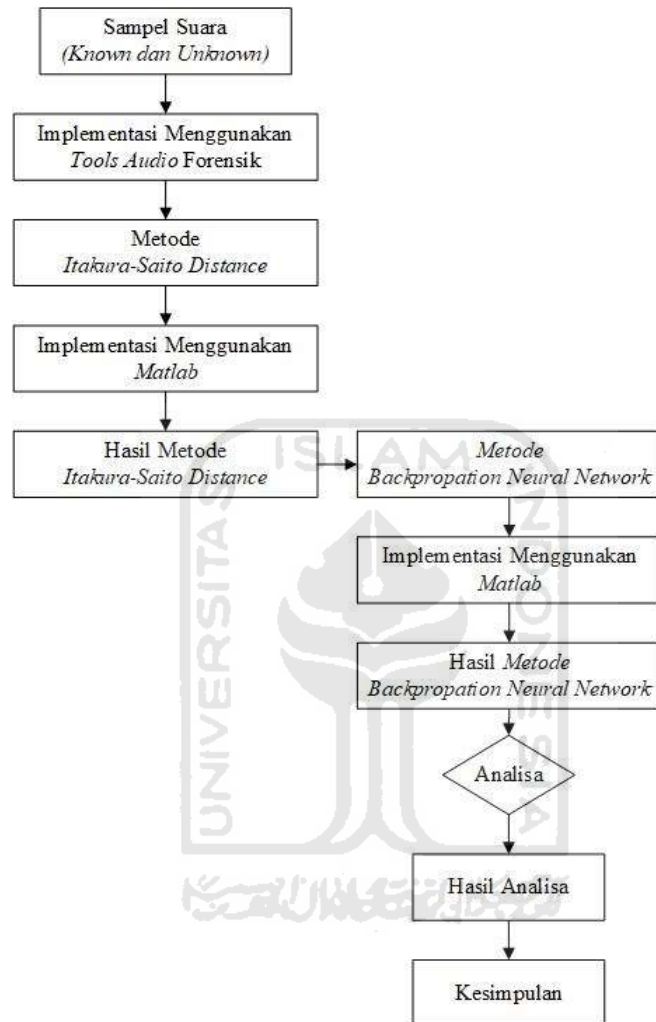
Jaringan dengan bobot 100 dan tingkat akurasi 0.1, maka 1000 pola *input* diharapkan akan mampu mengenali dengan benar 90% pola diantaranya.



## BAB 3

### Metodologi

Secara garis besar metodologi penelitian ini dapat dilihat pada Gambar 3. 1.



Gambar 3. 1 Alur metode penelitian

Metode penelitian ini berdasarkan *review* dari penelitian terkait sebelumnya dengan tujuan untuk mengembangkan pemikiran penelitian yang sudah ada, dimana dilakukan sebuah analisis untuk mengetahui tahapan-tahapan apa saja yang diterapkan dalam penggunaan metode yang sesuai dalam proses penanganan barang bukti *audio forensic*. Penelitian ini untuk menguji coba agar mendapatkan kesimpulan dari pemanfaatan metode apa yang tepat untuk penanganan barang bukti yang berupa rekaman *audio* atau suara. Penelitian ini memiliki enam tahapan diantaranya identifikasi masalah, tinjauan pustaka, konsep dasar *audio forensic*, perumusan metode *audio forensic*, analisis data, dan laporan yang kemudian

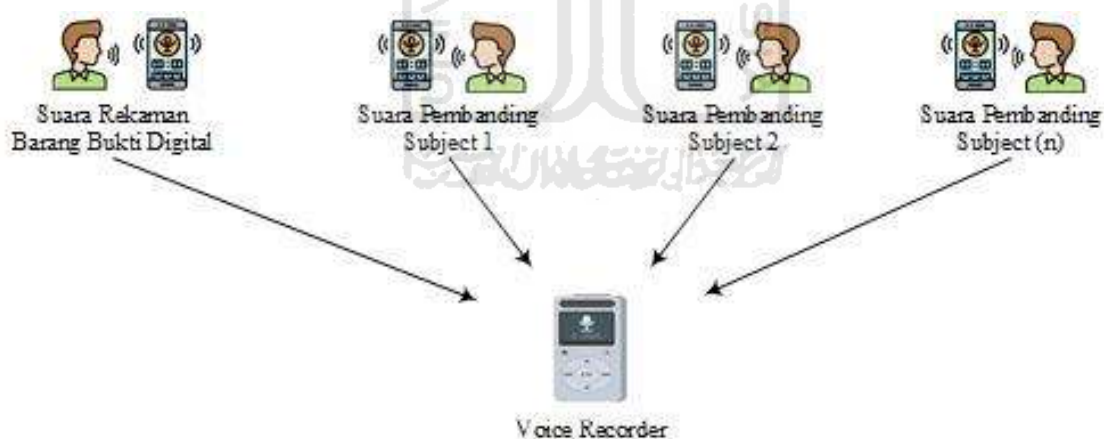
bisa ditarik kesimpulan berkaitan dengan hasil kombinasi metode untuk menunjang proses penanganan barang bukti rekaman *audio*.

### 3.1. Tinjauan Pustaka

Tinjauan pustaka dilakukan untuk mengumpulkan bahan-bahan referensi yang terkait dengan penelitian dari berbagai sumber diantaranya melalui buku, artikel jurnal, paper makalah, dan situs website yang terkait dengan metode *Itakura-Saito Distance* dan metode *Backpropagation Neural Network* yang dilakukan untuk perbandingan barang bukti rekaman suara.

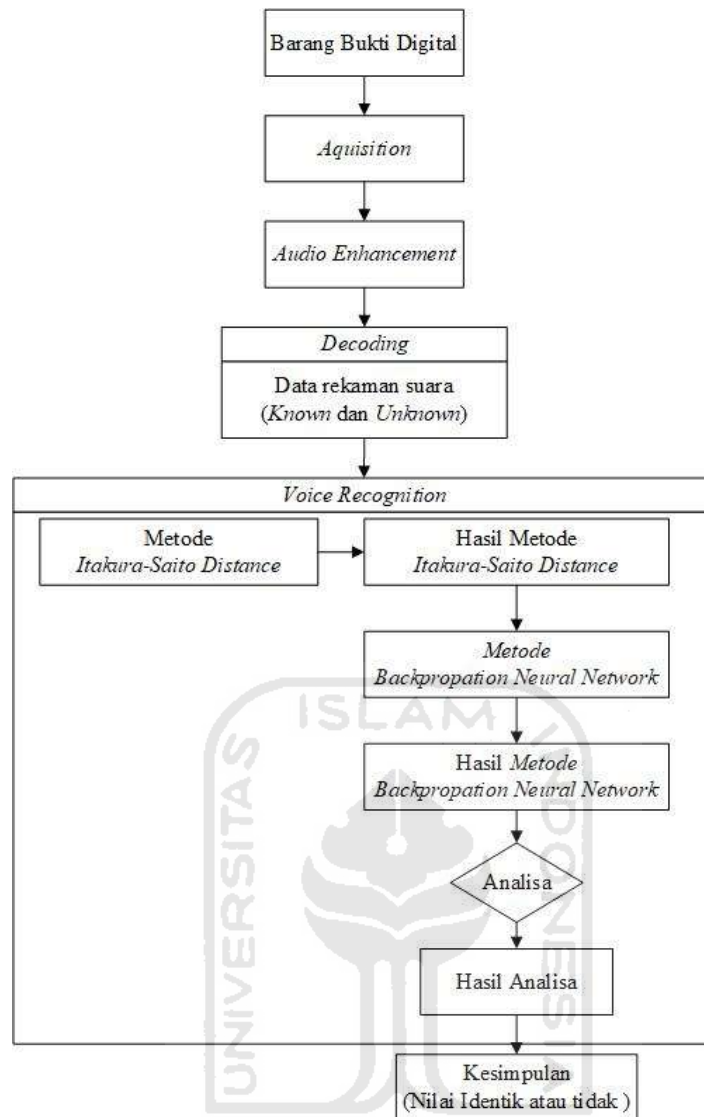
### 3.2. Teknik Pengambilan Sampel Data

Tahapan ini digunakan untuk memahami teknik pengambilan sampel rekaman suara dan prosedur-prosedur yang digunakan dan sesuai dengan *Standard Operating Procedure (SOP)* tentang analisis *audio forensic* dari *Digital Forensic Analysis Team (DFAT)*, sehingga barang bukti yang diperoleh harus bisa menjamin *Chain of Custody (CoC)*. Pengambilan data dilakukan menggunakan alat perekam suara dengan pengucapan kata-kata minimal 20 kata untuk dapat menunjukkan keidentikan suara antara *unknown* dengan *known* (pembanding) (Al-Azhar Nuh, 2011). Ilustrasi pengambilan sampel rekaman suara dapat dilihat pada Gambar 3. 2.



Gambar 3. 2 Ilustrasi pengambilan sampel rekaman suara

Pengambilan sampel suara diatas digunakan untuk proses implementasi beberapa metode yang akan dikombinasikan dan dianalisis untuk mengetahui tingkat keakuratan dalam melakukan identifikasi rekaman suara dengan mengacu pada rumusan aspek-aspek utama sesuai parameter yang digunakan oleh setiap metode. Skema proses penanganan barang bukti digital rekaman *audio* bisa pada Gambar 3. 3.



Gambar 3. 3 Prosedur penanganan barang bukti digital rekaman suara

Tahapan penanganan barang bukti *audio forensic* yang diperoleh berdasarkan skenario kasus dengan memiliki beberapa tahapan sesuai dengan SOP yaitu:

#### 1. *Acquisition*

Proses akuisisi dilakukan terhadap barang bukti elektronik yang berisikan rekaman suara asli (*unknown*) yang ditemukan ditempat kejadian perkara maupun barang bukti rekaman yang akan dijadikan sebagai rekaman suara pembanding (*known*). Berdasarkan skenario rekaman suara yang ditemukan dari proses akuisisi *audio recorder* akan menghasilkan *file DD image* sesuai dengan yang ditetapkan, dengan penemuan *file* yang berekstensi *.wav*. Berikut alur dari proses akuisisi dapat dilihat pada Gambar 3. 4.



Gambar 3. 4 Alur Proses Akuisisi Barang Bukti Digital

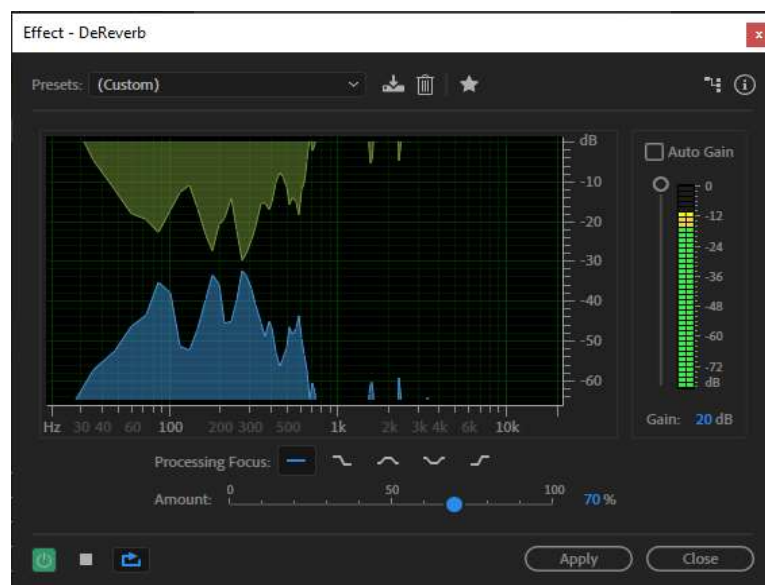
## 2. Audio Enhancement

Tahapan ini merupakan proses evaluasi rekaman suara yang ditemukan dengan memperdengarkan (*playback*) untuk mengetahui kualitas suaranya. Jika rekaman suara kualitas suara kurang baik mungkin dikarenakan adanya faktor *noise*, maka diperlukan proses untuk menghilangkan *noise* tersebut. Proses evaluasi suara yang terdapat *noise* mungkin akan mengganggu dalam proses identifikasi, salah satunya suara menggemema dalam ruangan atau hal lainnya. Penelitian ini menggunakan aplikasi Adobe Audition CC 2019. Berikut adalah ilustrasi dari proses *audio enhancement* dapat dilihat pada Gambar 3. 5.



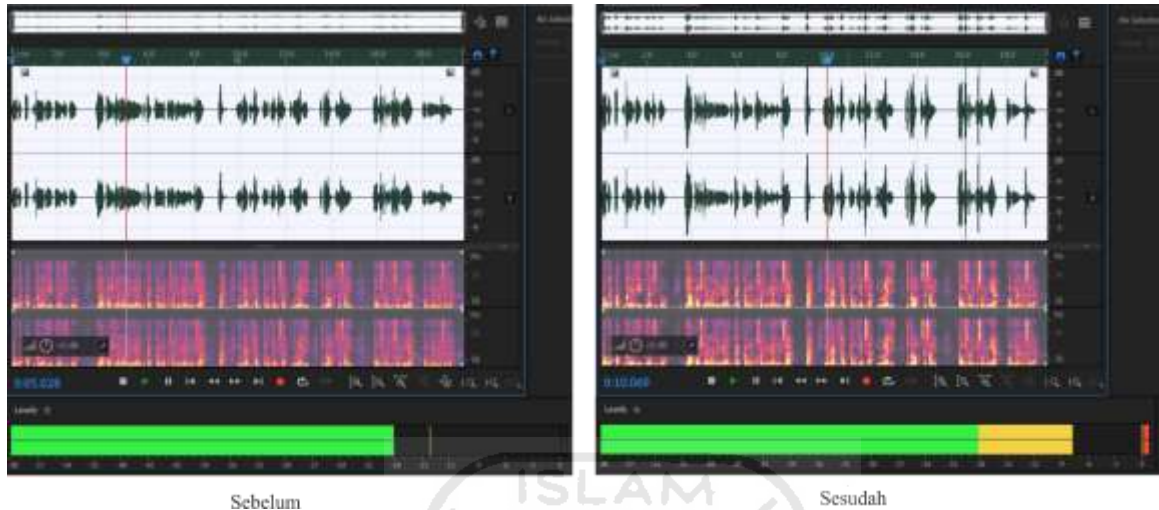
Gambar 3. 5 Ilustrasi proses *audio enhancement*

Proses penghilangan *noise* salah satunya bisa menggunakan *tools* bawaan dari Adobe Audition CC 2019 yaitu DeReverb yang terdapat pada menu *Effects-Noise Reduction/Restoration-DeReverb*. Berikut *tools* DeReverb dapat dilihat pada Gambar 3. 6.



Gambar 3. 6 Tampilan *effects DeReverb*

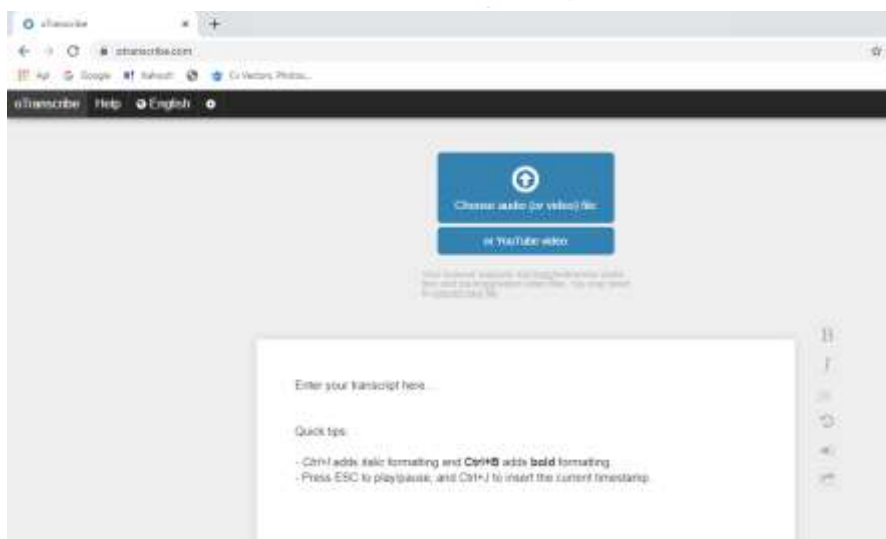
Pengaturan *DeReverb* tergantung dari kondisi suara menyesuaikan tingkat dan jenis *noise* yang ingin dihapus, sebagai contoh terdapat *tools processing focus* yaitu untuk memfokuskan frekuensi mana yang ingin dihapus. Setelah proses penghilangan *noise* diperoleh perbedaan grafik dan *spectrogram* yang dapat dilihat pada Gambar 3. 7.



Gambar 3. 7 Perbedaan hasil rekaman suara pada proses penghilangan *noise*

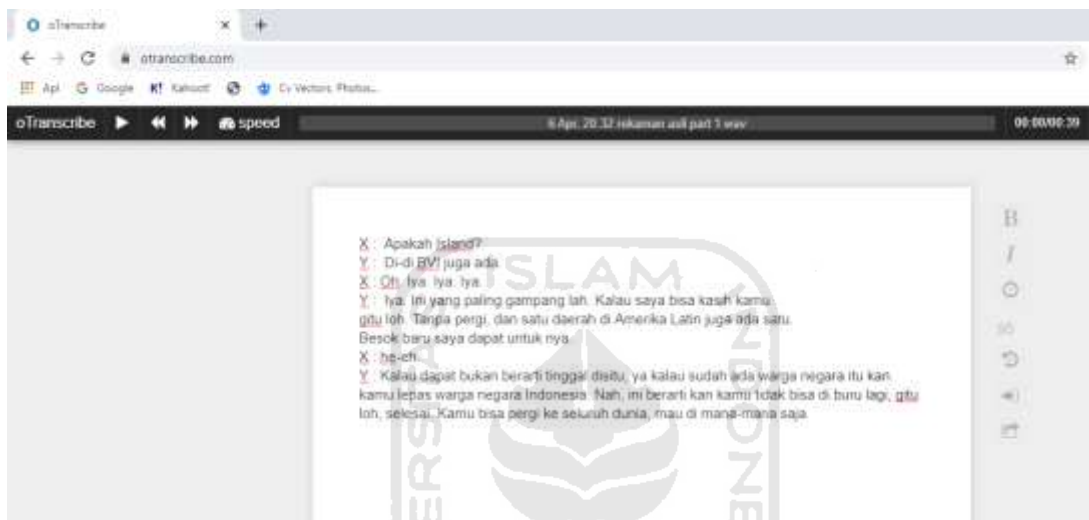
### 3. *Decoding*

Tahap ini adalah proses transkrip dari rekaman suara yang ditemukan, dan pembuatan transkrip ini dilakukan dengan tujuan membantu mendapatkan nilai akurasi yang lebih presisi terhadap hasil transkrip. Penelitian ini menggunakan proses transkrip secara online melalui situs <https://otranscribe.com/>. Proses memulai transkrip klik “Mulai Transkripsi” maka akan muncul halaman yang dapat dilihat pada Gambar 3. 8.



Gambar 3. 8 Tampilan halaman kerja oTranscribe

Transkrip hasil dimulai dari proses *input* dari rekaman suara kemudian melakukan transkrip, dan yang menjadi fokus yang perlu dianalisis terdapat pada kalimat “Kalau dapat bukan berarti tinggal disitu, ya kalau sudah ada warga negara itu kan kamu lepas warga negara Indonesia. Nah, ini berarti kan kamu tidak bisa di buru lagi, gitu loh, selesai. Kamu bisa pergi ke seluruh dunia, mau di mana-mana saja.” yang diucapkan oleh satu tersangka. Kalimat tersebut memiliki lebih dari 20 kata rekaman, maka dari potongan rekaman tersebut dapat dilakukan proses *audio forensic* yang sesuai dengan syarat standar SOP yang ditentukan. Berikut hasil dari transkrip di *oTranscribe* dapat dilihat pada Gambar 3. 9.



Gambar 3. 9 Hasil transkrip percakapan dari *audio sample* barang bukti

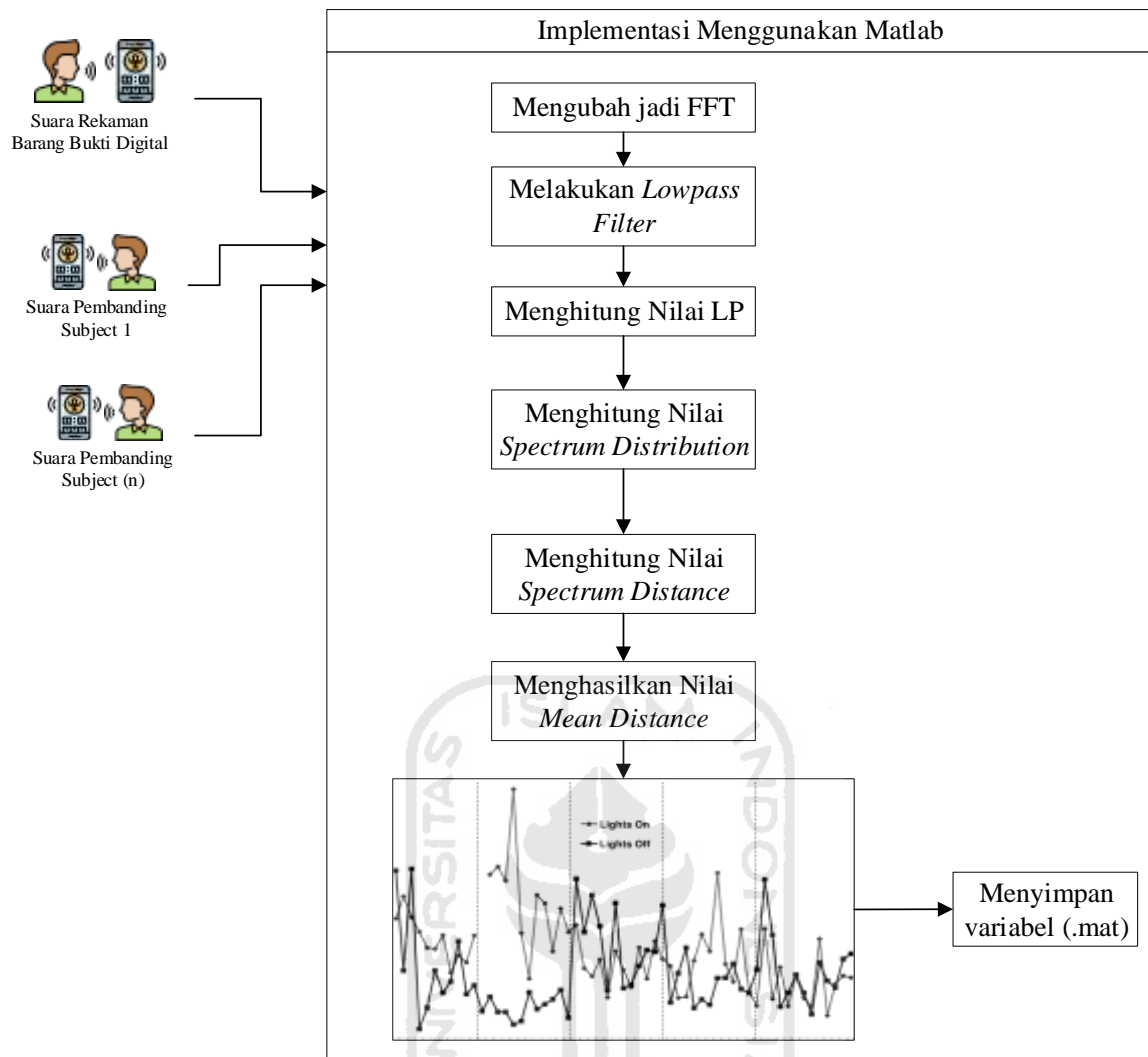
#### 4. Voice Recognition

Voice Recognition adalah teknik dimana proses analisis atau identifikasi apakah rekaman suara asli (*Unknown*) identik dengan rekaman suara pembanding (*Known*) yang didapatkan dari tahapan *sampling*. Analisis dilakukan menggunakan Matlab 2015b. Sesuai dengan ketentuan *audio forensic*, pada proses ini harus memiliki minimal 20 kata yang memiliki kesamaan antara rekaman suara barang bukti (*unknown*) dengan rekaman suara pembanding (*known*). Penelitian ini akan melakukan analisis rekaman suara pada proses *voice recognition* yaitu menggunakan Metode *Itakura-saito Distance* dan Metode *Backpropagation Neural Network*.

### 3.3. Implementasi Metode

Berikut skema proses implementasi metode *Itakura Saito Distance* dapat dilihat pada Gambar 3. 10.



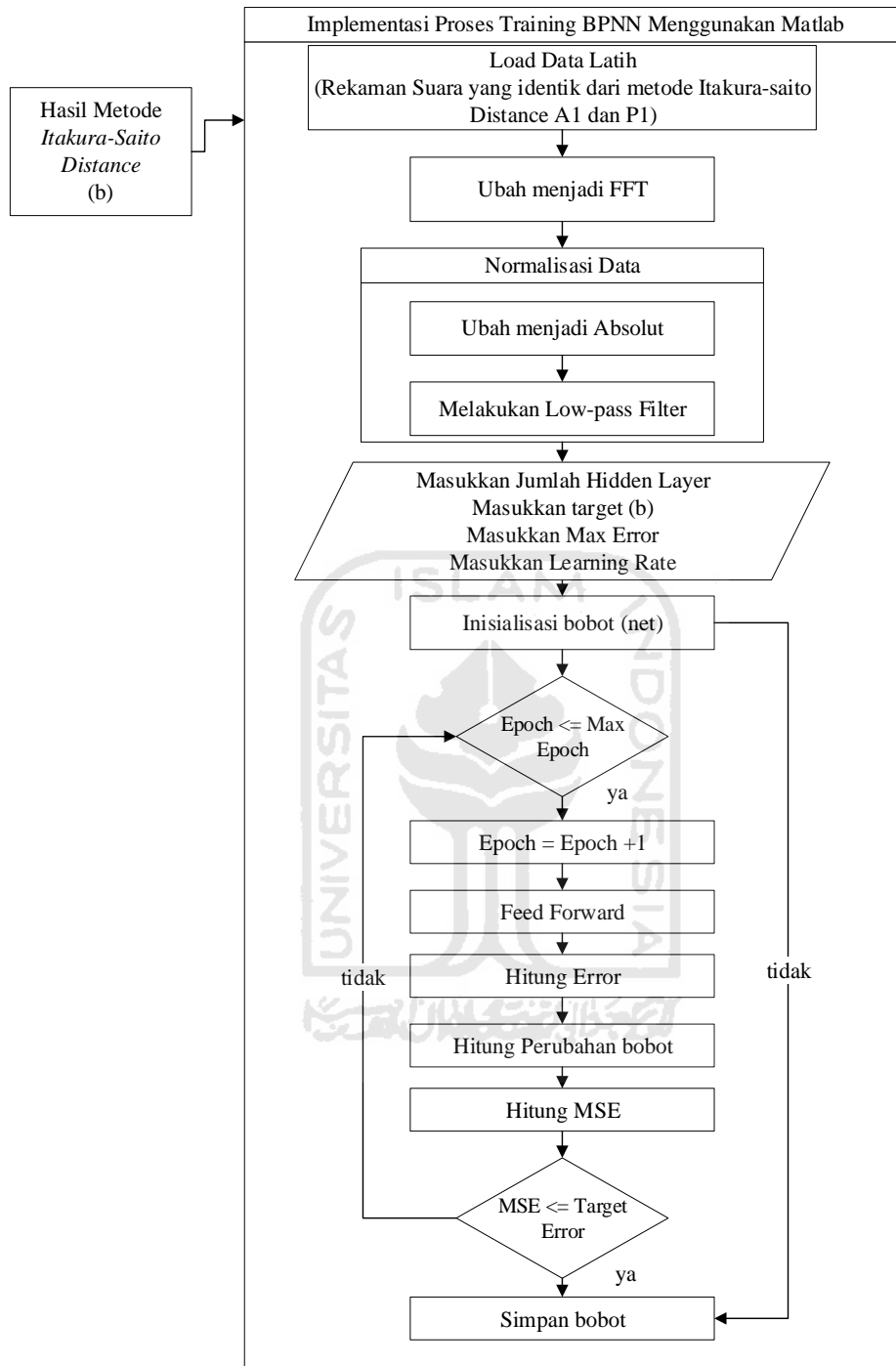


Gambar 3. 10 Tahap implementasi metode *Itakura-Saito Distance*

Metode *Itakura Saito Distance* memiliki beberapa tahapan dalam proses analisis yaitu hasil rekaman suara dari proses *audio enhancement* dengan format *.wav*. Rekaman suara menjadi *input* berupa sinyal analog. Sinyal analog diubah menjadi sinyal frekuensi (*spectrum*) dengan FFT. Selanjutnya melakukan tahap *Lowpass filter* untuk menghilangkan *high frequency noise* yang masih terdapat pada *audio enhancement*. Tahap selanjutnya mengevaluasi kualitas bicara berdasarkan *linear prediction* antara rekaman suara (A1) dan rekaman suara pembanding (P1) yang dianalisis dengan melakukan perhitungan LPC. Untuk menghitung LPC dibutuhkan rentang waktu tertentu (*frame*). Kemudian menghitung jarak *spectrum* antara dua rekaman suara.

Selanjutnya akan diperoleh nilai rentan jarak untuk menentukan rekaman suara identik atau tidak dengan rekaman suara pembanding, kemudian nilai yang diperoleh menjadi ekstraksi ciri (nilai target) dan dianalisa dengan metode *Backpropagation Neural Network*. Metode tersebut diharapkan dapat memperoleh hasil perbandingan suara menjadi

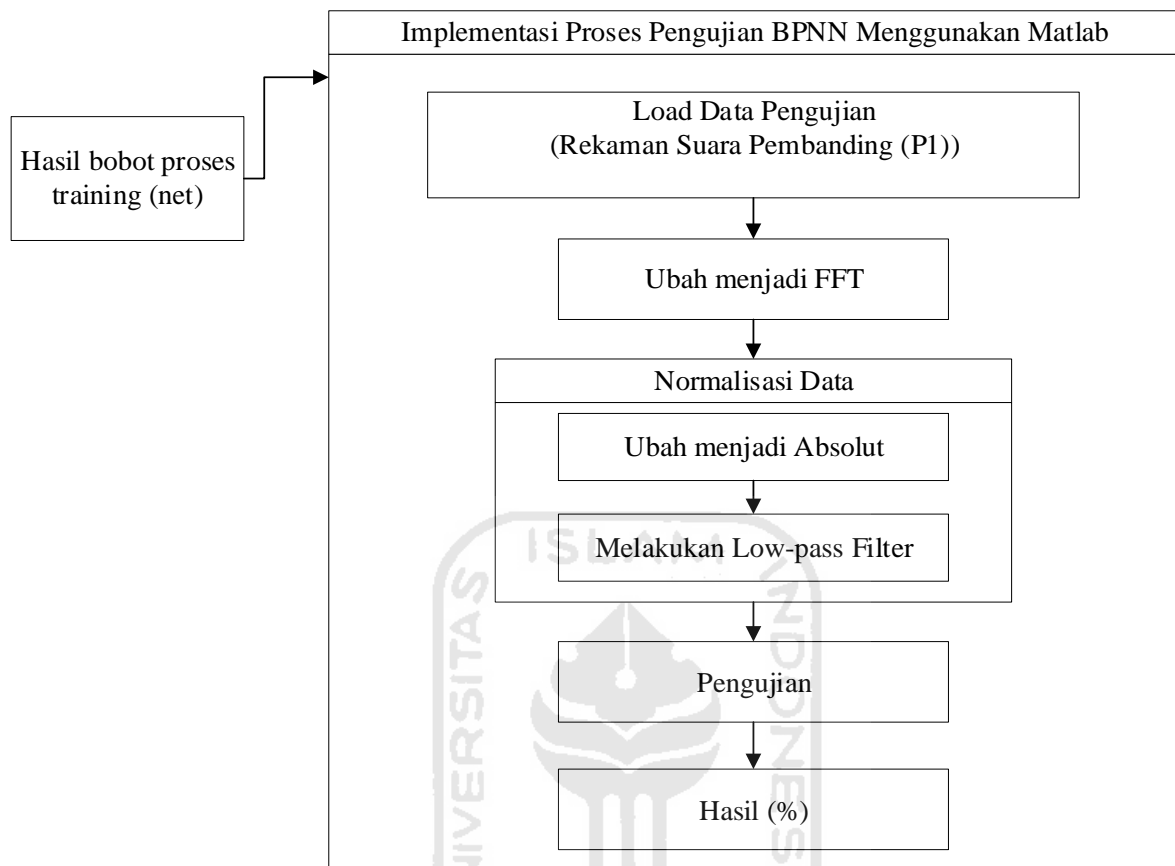
akurat. Skema implementasi metode *Backpropagation Neural Network* dapat dilihat pada Gambar 3. 11.



Gambar 3. 11 Implementasi metode *Backpropagation Neural Network*

Pengecekan *error* dengan cara menghitung nilai *Mean Square Error* (MSE), apabila  $MSE_{input} \leq MSE$  maka proses dihentikan, jika tidak maka proses akan diteruskan untuk membaca data pertama lagi ( $i = 1$ ) hingga syarat  $MSE_{hitung} \leq MSE_{input}$  terpenuhi.

Selanjutnya pada tahap pengujian, dimana akan dilakukan perhitungan semua *hidden layer* dan *output layer*. Proses pengujian dapat dilihat pada Gambar 3. 12.



Gambar 3. 12 Implementasi proses *training* metode *Backpropagation Neural Network*

### 3.4. Analisis Hasil Data

Tahap ini adalah proses analisis data dari hasil pengujian metode *Itakura-Saito Distance* dan metode *Backpropagation Neural Network* yang menghasilkan kesimpulan bahwa metode tersebut dapat menghasilkan nilai keakuratan dalam perbandingan rekaman suara.

Tabel 3. 1 Analisis Implementasi Metode

Metode	Pengujian 1	Pengujian 2	Pengujian (n)	Hasil Pengujian			
Metode 1	R <sub>1</sub>	S <sub>11</sub>	S <sub>21</sub>	S <sub>m1</sub>			
		S <sub>12</sub>	S <sub>22</sub>	S <sub>m2</sub>			
		S <sub>1n</sub>	S <sub>2n</sub>	S <sub>mn</sub>			
Metode 2	R <sub>2</sub>	S <sub>21</sub>	S <sub>21</sub>	S <sub>m1</sub>			
		S <sub>22</sub>	S <sub>22</sub>	S <sub>m2</sub>			
		S <sub>2n</sub>	S <sub>2n</sub>	S <sub>mn</sub>			
Metode n	R <sub>n</sub>	S <sub>1n</sub>	R <sub>n</sub>	S <sub>2n</sub>	R <sub>m</sub>	S <sub>mn</sub>	

Tabel 3. 2 Hasil evaluasi metode

Subyek Suara	Total Nilai	Keterangan
R1 dan S1	Nilai 1	Identik dengan rekaman (n)
R1 dan S2	Nilai 2	
R(n) dan S(n)	Nilai (n)	

Keterangan:

R: Rekaman Bukti Digital

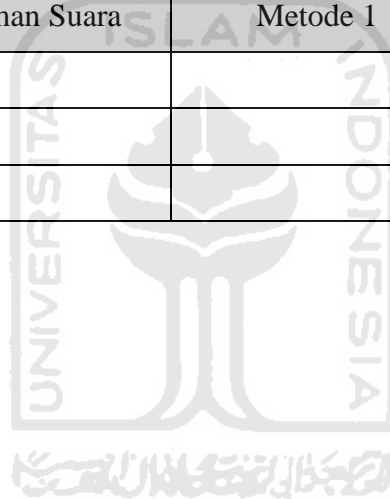
S: Subject (Rekaman Suara Pembanding)

### 3.5. Laporan

Tahap ini berisi kesimpulan hasil pengujian dan analisis metode *Itakura-Saito Distance* dan metode *Backpropagation Neural Network*.

Tabel 3. 3 Karakter hasil analisis metode

No	Jenis Karakter Rekaman Suara	Metode 1	Metode 2
1	Variabel <sub>1</sub>		
2	Variabel <sub>2</sub>		
n	Variabel <sub>n</sub>		



## BAB 4

### Hasil dan Pembahasan

#### 4.1. Analisis Kebutuhan

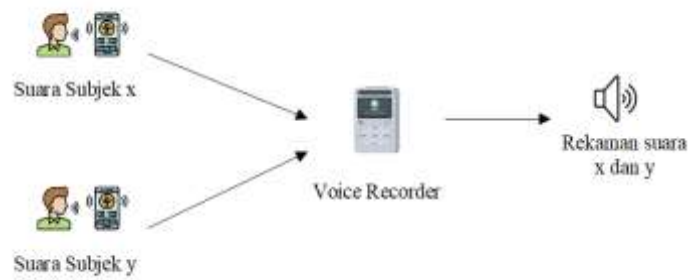
Melakukan uji coba terhadap implementasi metode yang ada diperlukan kebutuhan beberapa *software* dan *hardware* untuk mendukung proses analisis terhadap metode yang dikembangkan, adapun kebutuhannya sebagai berikut:

1. *Software*
  - a. Windows 10 Pro 64 Bit
  - b. Adobe Audition 2019
  - c. Matlab 2015b
2. *Hardware*
  - a. Komputer
    - 1) *Processor Intel® Core™ i7-7700*
    - 2) *RAM 8 GB*
    - 3) *Harddisk 500 GB*
  - b. *Smartphone*
    - 1) *Device Model Xiaomi Redmi 5 Plus*
    - 2) *Storage 32 GB*
    - 3) *RAM 3 GB*
  - c. *Earphone/Headset*



#### 4.2. Skenario Kasus

Melakukan analisis terhadap implementasi metode yang ada diperlukan skenario kasus secara eksperimen dengan melibatkan rekaman suara asli (*unknown*) dan rekaman suara pembandingan (*known*). Ilustrasi kasus dapat dilihat Gambar 4. 1.



Gambar 4. 1 Ilustrasi Skenario Kasus

Diskripsi kasusnya dimana X merupakan orang dalam penyidikan KPK dalam kasus suap panitera. Sehingga X ingin mengubah kewarganegaraan untuk melarikan diri ke luar negeri dalam menghindari proses hukum di KPK. Pada tanggal 06 April 2020 X bertemu dengan Y yang disuruh untuk membantu dalam aksi tersebut, tetapi percakapan tersebut terdengar oleh salah karyawan yang berada disekitar ruangan X dan berinisiatif merekam percakapan tersebut menggunakan perangkat *smartphone* secara diam-diam.

Pada tanggal 08 April 2020 karyawan tersebut memberanikan memberikan rekaman yang dia peroleh kepada polisi, sehingga X dan Y dipanggil untuk mengklarifikasi perbuatan tersebut. Barang bukti yang diberikan berupa barang bukti elektronik yaitu *smartphone* sebagai alat rekaman suara yang hasil dari rekaman tersebut adalah dugaan membantu melarikan seseorang yang berstatus tersangka KPK.

#### 4.3. Analisis Metode

Eksperimen skenario kasus diperoleh sebuah barang bukti rekaman suara yang akan dianalisis. Melalui tahapan *decoding* diperoleh transkrip secara lengkap isi percakapan rekaman suara yang terduga sebagai tersangka. Selanjutnya transkrip rekaman dianalisis untuk menentukan bagian percakapan yang dapat dilihat pada Tabel 4. 1.

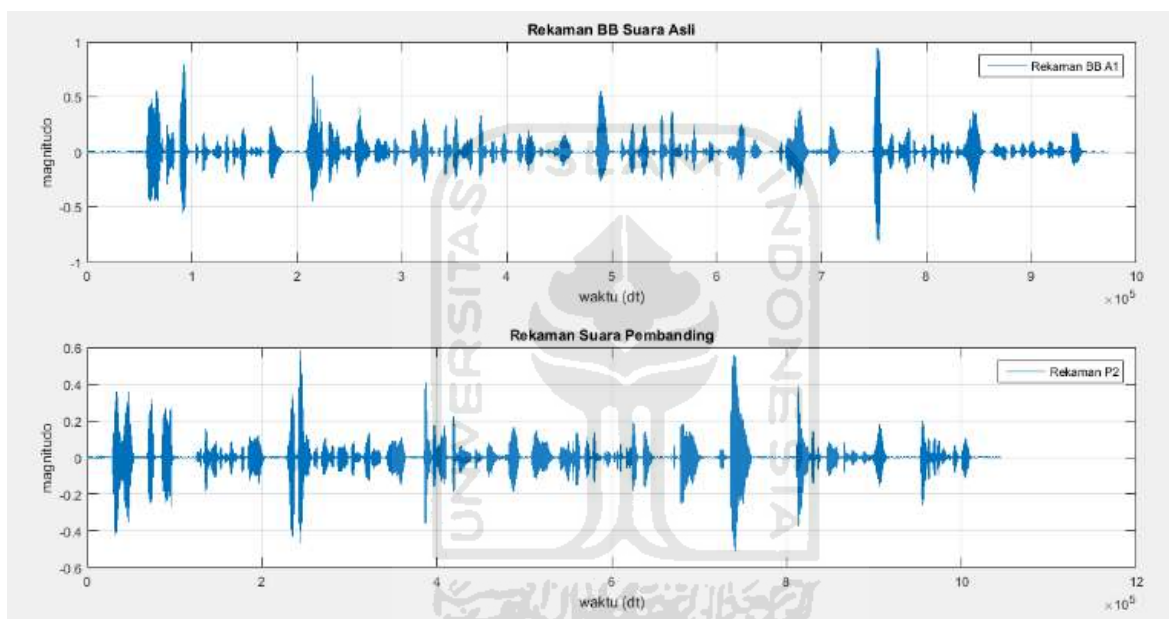
Tabel 4. 1 Penentuan data analisis

Subjek Rekaman	Kalimat	Jumlah
Rekaman suara <i>unknown</i> ( $A_n$ )	“Kalau dapat bukan berarti tinggal disitu, ya kalau sudah ada warga negara itu kan kamu lepas warga negara Indonesia. Nah, ini berarti kan kamu tidak bisa di buru lagi, gitu loh, selesai. Kamu bisa pergi ke seluruh dunia, mau di mana-mana saja”	36 kata/subjek
Rekaman suara <i>known</i> ( $P_n$ )		

Setelah menentukan kalimat panjang yang akan dijadikan analisis, selanjutnya mencari rekaman suara dari beberapa pembanding. Dimana untuk memperkuat asumsi dari implementasi metode yang digunakan, penelitian ini menggunakan 4 rekaman suara *known* dan rekaman suara *unknown*. Implementasi metode dilakukan pada tahapan *voice recognition* dengan menggunakan aplikasi Matlab 2015b yaitu:

#### 4.3.1. Itakura-Saito Distance

*Itakura Saito Distance* dapat mencerminkan perbedaan *spectrum amplitude* waktu dengan jarak dari dua rekaman suara, sehingga dapat mendistorsi sinyal ucapan dengan baik. Rekaman suara yang sudah dilakukan tahap *audio enhancement* akan menjadi *input* berupa sinyal analog seperti terlihat pada Gambar 4. 2.



Gambar 4. 2 Sinyal analog Rekaman A1 dan P2

Sinyal analog tersebut diubah kedalam bentuk sinyal frekuensi menggunakan algoritma FFT. Untuk memperbaiki kualitas suara yang masih terdapat pada *audio enhancement* dilakukan proses *Lowpass Filter*. Berikut proses untuk melakukan *low-pass filter* pada rekaman suara A1 dapat dilihat pada Gambar 4. 3.

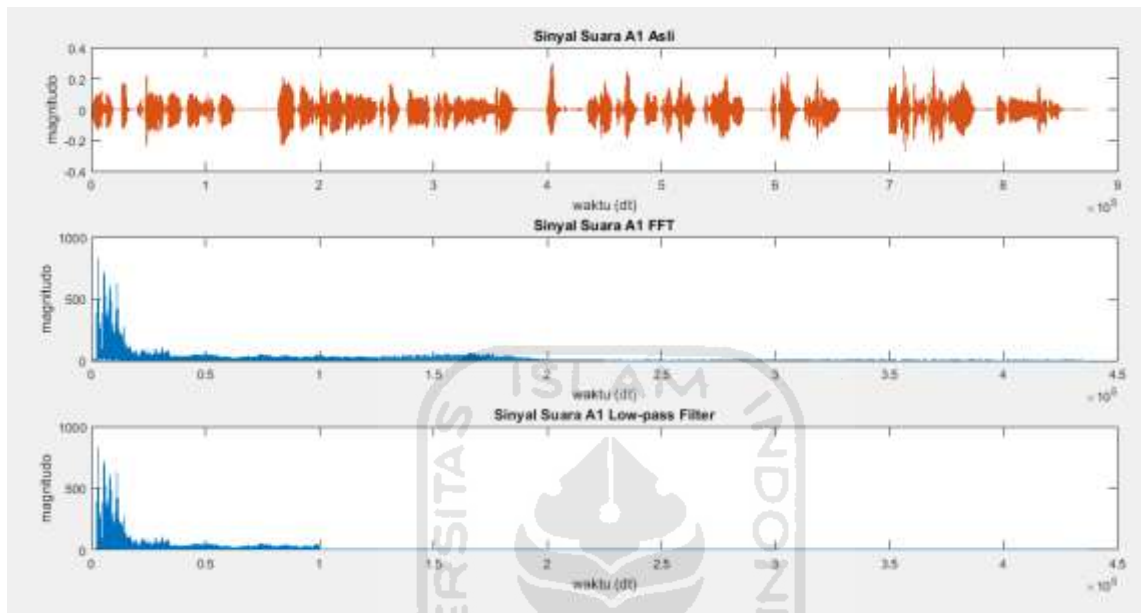
```

2 %% Sound Analysis Thesis 18917107
3 - pj = 100000; %%batas dari panjang waktu
4 - [al, fs1] = audioread('1_A1.wav'); %%input suara
5 - al = al(1:end,1)+al(1:end,2); %%normalisasi
6 - al = al/2;
7 - al = LowPassFilter(al, 0, pj, ''); %%fungsi lowpass filter
8

```

Gambar 4. 3 Proses melakukan low-pass filter

Gambar 4. 3 dijelaskan bahwa *input* adalah suara yang dipanggil dengan kode *audioread*, dimana dari hasilnya dinormalisasi menjadi satu *channel*. Kode “*LowPassFilter*” merupakan *function* tersendiri dari proses *low-pass filter*. Penelitian ini menggunakan nilai batas panjang waktu (*frame*) sebesar 100000 dari jumlah keseluruhan panjang waktu dari setiap suara itu sendiri. Berikut hasil dari sampel satu suara “A1” yang dilakukan proses FFT dan *Lowpass Filter* dapat dilihat pada Gambar 4. 4.



Gambar 4. 4 Hasil dari memperbaiki kualitas suara dengan *low-pass filter*

Pada Gambar 4. 4, *plot Lowpass Filter* terlihat perbedaan pengurangan frekuensi menjadi 0 dari rentang waktu 1 ke seterusnya. Selanjutnya mengevaluasi kualitas bicara berdasarkan *linear prediction* antara rekaman suara (A1) dan rekaman suara pembanding (P1) yang dianalisis dengan melakukan perhitungan dengan *function* LPC. Berikut proses perhitungan LPC dapat dilihat pada Gambar 4. 5.

```

% Compute autocorrelation vector or matrix
X = fft(x,2^nextpow2(2*size(x,1)-1));
R = ifft(abs(X).^2);
R = R./m; % Biased autocorrelation estimate

```

Gambar 4. 5 Perhitungan LPC

Proses pemanggilan *function* LPC untuk rekaman suara (A1) dengan variabel *y1* dan rekaman suara (P1) dengan variabel *y2* dapat dilihat pada Gambar 4. 6.



```

%% Analisa LPC
y1 = lpc (a1,100000);
y1 = abs(fft(y1));
y2 = lpc (p2,100000);
y2 = abs(fft(y2));

y1=y1';
y2=y2';

```

Gambar 4. 6 Perintah pemanggilan *function* LPC

Perintah pada Gambar 4. 6 bertujuan untuk menghitung dan mengkalkulasi nilai LPC dengan rentang panjang waktu 100000 (*frame*). Berikut tampilan hasil LPC dari *y1* jika memproses 25 *frame* dapat dilihat pada Gambar 4. 7.

```

>> y1 = lpc (a1,25);
>> y1

y1 =

Columns 1 through 12
 1.0000  -4.7036  10.7873  -14.9471  12.1820  -3.6969  -3.1218  2.4219  2.2223  -3.3022  -0.3633  2.7681
 1.0000  -4.7036  10.7873  -14.9471  12.1820  -3.6969  -3.1218  2.4219  2.2223  -3.3022  -0.3633  2.7681

Columns 13 through 24
 -0.1750  -3.5106  2.8969  0.7381  -2.4756  0.3784  2.2265  -2.0852  -0.0917  1.6456  -1.5322  0.6900
 -0.1750  -3.5106  2.8969  0.7381  -2.4756  0.3784  2.2265  -2.0852  -0.0917  1.6456  -1.5322  0.6900

Columns 25 through 26
 -0.1365  0.0019
 -0.1365  0.0019

```

Gambar 4. 7 Hasil nilai LPC *y1* dari rekaman suara A1

Selanjutnya menentukan jarak *spectrum* antara rekaman suara A1 dengan rekaman suara P1 menggunakan metode *itakura-saito distance*. Hasil kalkulasi nilai LPC kemudian diproses dengan beberapa *function* seperti *lpcar2ra* digunakan untuk memfilter autokorelasi *inverse filter*, *lpcar2rr* untuk mengubah koefisien *autoregressive* menjadi koefisien autokorelasi, dan *lpcar2rf* untuk mengonversi koefisien *autoregressive* menjadi koefisien refleksi AR=RF. Menjadi pokok adalah *function lpcar2ra* yang dapat dilihat pada Gambar 4. 8.

```

[nf,pl]=size(ar);
ra=zeros(nf,pl);
for i=1:pl
    ra(:,i)=sum(ar(:,1:pl+1-i)).
end

```

Gambar 4. 8 Perintah mengkalkulasi LPC dengan *function lpcar2ra*

Kemudian dilanjutkan dengan proses metode *Itakura-Saito Distance*, dengan *function distisar* yang dapat dilihat pada Gambar 4. 9.

```

71 - [nf1,p1]=size(ar1);
72 - nf2=size(ar2,1);
73 - m2=lpcar2ra(ar2);
74 - m2(:,1)=m2(:,1)*0.5;
75 - if nargin<3 || isempty(mode) mode='0'; end
76 - if any(mode=='d') || (mode~='x' & nf1==nf2)
77 -     nx=min(nf1,nf2);
78 -     d=2*sum(lpcar2rr(ar1(1:nx,:)).*m2(1:nx,:),2)-log((ar2(1:nx,1)./ar1(1:nx,1)).^2)-1;
79 - else
80 -     d=2*lpcar2rr(ar1)*m2'-log((ar1(:,1).^(-1)*ar2(:,1)).^2)-1;
81 - end

```

Gambar 4. 9 Menghitung nilai jarak pada metode *Itakura-Saito Distance*

Sedangkan untuk menjalankan proses analisis *function* diatas dapat menggunakan perintah seperti terlihat pada Gambar 4. 10.

```

22 - %% Proses Itakura Saito Distance
23 - b1 = distisar (y1,y2,'d');

```

Gambar 4. 10 Perintah menjalankan *function Itakura-Saito Distance*

Dari proses perhitungan *itakura-saito distance* dapat diperoleh akan menghasilkan nilai *formant* pada Gambar 4. 11.

```

Column 99937 through 99948
  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000

Column 99949 through 99960
  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000

Column 99961 through 99972
  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000

Column 99973 through 99984
  1.1000  1.1000  0.9489  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000

Column 99985 through 99996
  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000  1.1000

Column 99997 through 100001
  1.1000  1.1000  1.1000  1.1000  1.1000

```

Gambar 4. 11 Hasil nilai *spectrum (formant)* dari proses *itakura-saito distance*

Selanjutnya nilai yang terdapat pada Gambar 4. 11 dianalisis untuk mencari tingkat akurasi identik atau tidak dari rekaman suara A1 dan rekaman suara P1. Dimana jika suara nilai *formant* mendekati 0 maka suara tersebut identik, dan jika mendekati 1 suara tersebut tidak identik. Berikut perintah pencarian suara identik dapat dilihat pada Gambar 4. 12.

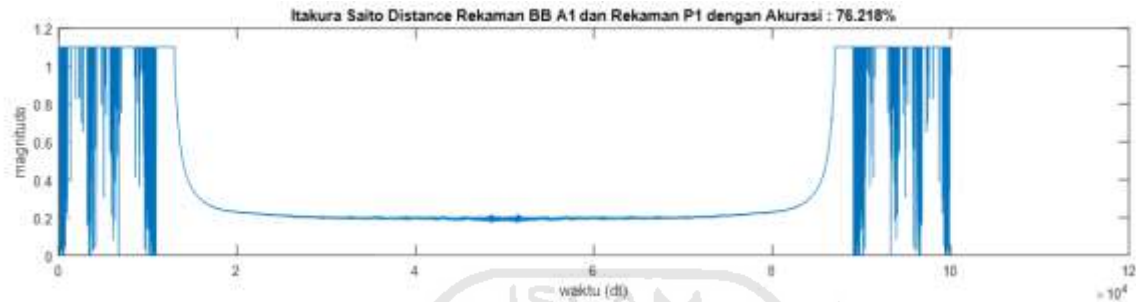
```

25 - b1(b1<=0) = 0;
26 - b1(b1>=1) = 1.1;
27 - acc = 100*(length(b1(b1<=0.9)))/100000);

```

Gambar 4. 12 Perintah pencarian akurasi suara identik

Sehingga rekaman suara A1 dan rekaman suara P1 memiliki tingkat akurasi sebesar 76,218 % dengan 100000 *frame* dari jumlah keseluruhan *frame*. Hasil tersebut dapat dilihat dari visualisasi yang terdapat pada Gambar 4. 13.



Gambar 4. 13 Hasil Visualisasi *Itakura-Saito Distance*

Untuk melihat keseluruhan hasil maka perlu dilakukan analisis keseluruhan dari empat rekaman suara asli (*unknown*) dari empat responden dan empat rekaman suara pembanding (*known*) dari empat responden. Berikut hasil dari analisis keseluruhan dari rekaman suara.

Tabel 4. 2 Hasil evaluasi perbandingan nilai itakura-saito distance pada Rekaman A1 dan Rekaman P1, P2, P3, P4

Subjek Suara	Total Nilai	Presentase Akurasi	Keterangan
A1 dan P1	100000	76,218%	Rekaman A1 Identik dengan Rekaman P3
A1 dan P2	100000	76,423%	
A1 dan P3	100000	93,861%	
A1 dan P4	100000	77,317%	

Pada Tabel 4. 2 merupakan hasil evaluasi dari metode *itakura-saito distance* dengan keseluruhan presentase dari semua rekaman suara A1 dan rekaman suara pembanding P3 memiliki nilai yang mendekati 0 sebesar 93,861% dari total nilai yang diproses 100000 *frame*. Dengan tingkat akurasi sebesar itu ada kemungkinan rekaman suara A1 identik dengan suara P3.

Tabel 4. 3 Hasil evaluasi perbandingan nilai *itakura-saito distance* pada Rekaman A2 dan Rekaman P1, P2, P3, P4

Subjek Suara	Total Nilai	Presentase Akurasi	Keterangan
A2 dan P1	100000	76,667%	Rekaman A2 Identik dengan Rekaman P2
A2 dan P2	100000	89,312%	
A2 dan P3	100000	68,535%	
A2 dan P4	100000	87,459%	

Pada Tabel 4. 3 merupakan hasil evaluasi dari metode *itakura-saito distance* dengan keseluruhan presentase dari semua rekaman suara A2 dan rekaman suara pembanding P2 memiliki nilai yang mendekati 0 sebesar 89,312% dari total nilai yang diproses 100000 *frame*. Dengan tingkat akurasi sebesar itu ada kemungkinan rekaman suara A2 identik dengan suara P2.

Tabel 4. 4 Hasil evaluasi perbandingan nilai *itakura-saito distance* pada Rekaman A3 dan Rekaman P1, P2, P3, P4

Subjek Suara	Total Nilai	Presentase Akurasi	Keterangan
A3 dan P1	100000	84,805%	Rekaman A3 Identik dengan Rekaman P4
A3 dan P2	100000	92,812%	
A3 dan P3	100000	73,030%	
A3 dan P4	100000	95,952%	

Pada Tabel 4. 4 merupakan hasil evaluasi dari metode *itakura-saito distance* dengan keseluruhan presentase dari semua rekaman suara A3 dan rekaman suara pembanding P4 memiliki nilai yang mendekati 0 sebesar 93,861% dari total nilai yang diproses 100000 *frame*. Dengan tingkat akurasi sebesar itu ada kemungkinan rekaman suara A3 identik dengan suara P4.

Tabel 4. 5 Hasil evaluasi perbandingan nilai *itakura-saito distance* pada Rekaman A4 dan Rekaman P1, P2, P3, P4

Subjek Suara	Total Nilai	Presentase Akurasi	Keterangan
A4 dan P1	100000	94,011%	Rekaman A4 Identik dengan Rekaman P1
A4 dan P2	100000	85,724%	
A4 dan P3	100000	83,763%	
A4 dan P4	100000	87,639%	

Pada Tabel 4. 5 merupakan hasil evaluasi dari metode *itakura-saito distance* dengan keseluruhan presentase dari semua rekaman suara A4 dan rekaman suara pembanding P1 memiliki nilai yang mendekati 0 sebesar 93,861% dari total nilai yang diproses 100000 *frame*. Dengan tingkat akurasi sebesar itu ada kemungkinan rekaman suara A4 identik dengan suara P1.

#### 4.3.2. *Backpropagation Neural Network*

Berdasarkan hasil dari metode *Itakura Saito Distance* selanjutnya dianalisis kembali untuk mendapatkan tingkat akurasi yang lebih tinggi lagi. Sebagai contoh suara A1 dan P3 identik yang disimpan dengan variabel b3 dengan format ekstensi *mat*. Rekaman suara yang identik dari masing-masing perbandingan akan jadi acuan untuk meningkatkan akurasi. Proses penyimpanan dapat dilihat pada Gambar 4. 14.

```
save('D:\Thesis\Fix Analisis File\File Analisis Tesis\b3')
```

Gambar 4. 14 Save hasil *itakura-saito distance* dengan ekstensi ".mat"

Berikut proses dari metode *Backpropagation Neural Network* yaitu:

##### 1. *Proses training Backpropagation Neural Network*

Proses pertama dalam metode *Backpropagation Neural Network* yaitu *load data* dua rekaman suara yang diduga identik dan akan dijadikan data *training*. Sama halnya dengan proses yang ada di metode *Itakura Saito Distance*, data yang akan diuji harus dinormalisasi terlebih dahulu dengan mengubah sinyal ke domain frekuensi (FFT) dan memperbaiki kualitas suara memakai *low-pass filter*. Kemudian diperoleh hasil nilai  $x_1$  dan  $y_1$  dilakukan *invers* matriks yang nilai ini akan dijadikan nilai *input*. Proses *load data* dan normalisasi data dapat dilihat pada Gambar 4. 15.

```
1 - clear, clc, close all
2 - %% awal
3 - load('D:\Thesis\Fix Analisis File\File Analisis Tesis\b3.mat')
4 -
5 - %% Load data
6 - [x,fs] = audioread('1_A1.wav');
7 - HsFFT = fft(x(:,1));
8 - x1 = abs(HsFFT(1:100000,1))';
9 -
10 - [y,fs2] = audioread('1_P3.wav');
11 - HsFFT2 = fft(y(:,1));
12 - y1 = abs(HsFFT2(1:100000,1))';
```

Gambar 4. 15 Proses *load data* dan normalisasi

Hasil data  $x_1$  dan  $y_1$  kemudian digabungkan menjadi satu matriks, karena dimensinya berbeda kemudian dilakukan *invers* untuk dijadikan dalam dimensi matriks yang sama. Hal itu dilakukan supaya dimensi antara *load* data dengan data target mempunyai matriks yang sama. Proses *invers* dapat dilihat pada Gambar 4. 16.

```
%%  
%%pengubahan matriks invers  
xy =[x1' y1'];
```

Gambar 4. 16 Proses *invers* matriks

Tahap selanjutnya adalah memasukkan parameter untuk proses melakukan *training*. Penentuan nilai parameter berikut dilakukan secara eksperimen karena belum ada metode yang baku dalam penentuan parameter tersebut. Berikut parameter dalam proses *training*, yaitu:

- Panjang nilai dari rekaman suara A1 dan P1 sebagai *input layer*
- 60 *neuron* dalam 1 *hidden layer*
- 1 *output layer*
- Fungsi aktivasi menggunakan *sigmoid bipolar* (*tansig*)
- Menggunakan metode penurunan *gradient*, variasi *learning rate*, dan momentum
- Batas MSE 0,000001
- Batas *learning rate* 0,05
- Bobot diinisialisasikan secara acak

Proses dalam *input* parameter *training* dapat dilihat pada Gambar 4. 17.

```
net.trainParam.show = 50;  
net.trainParam.lr = 0.05;  
net.trainParam.goal = 1e-6;
```

Gambar 4. 17 *Input* parameter *training*

Selanjutnya adalah proses inialisasi bobot, dimana proses ini tergantung *epoch time* yang diberikan. *Epoch time* dalam penelitian ini sesuai dengan bawaan yaitu 1000 *epoch time*. Untuk membantu insialisasi bobot pada tetode *backpropagation neural network* menggunakan *toolbox* bawaan dari Matlab 2015b yaitu *newff*. Berikut merupakan arsitektur dari proses penentuan bobot dengan *newff* dapat dilihat pada Gambar 4. 18.

```

% Architecture
N1 = length(s)+1;
net = network;
net.numInputs = 1;
net.numLayers = N1;
net.biasConnect = ones(N1,1);
net.inputConnect(1,1) = 1;
[j,i] = meshgrid(1:N1,1:N1);
net.layerConnect = (j == (i-1));
net.outputConnect(N1) = 1;

```

Gambar 4. 18 Arsitektur *newff*

Berikut perintah pemanggilan *newff* dalam melakukan inialisasi bobot dan proses *backpropagation neural network* dari parameter yang dimasukkan. Perintah dapat dilihat pada Gambar 4. 19.

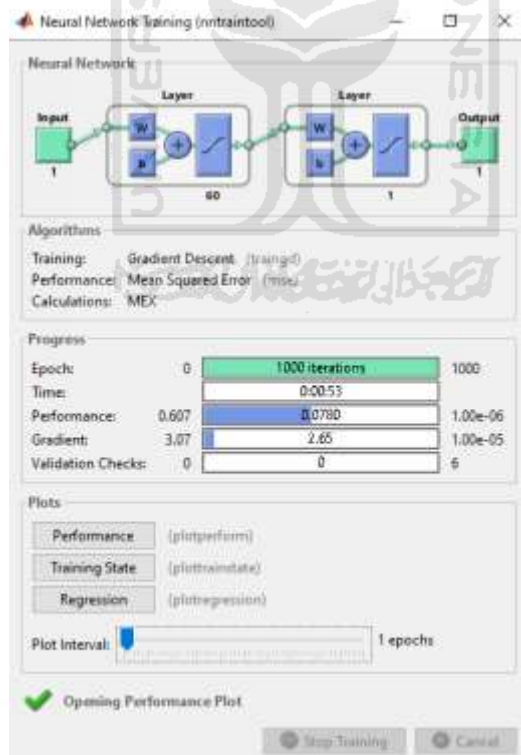
```

net = newff(minmax(x1), [60,1], {'tansig', 'tansig'}, 'traingd');
[net, tr] = train(net, xy(1:end,1)', b3(1:length(b3)-1,1)');

```

Gambar 4. 19 Perintah proses *Backpropagation Neural Network*

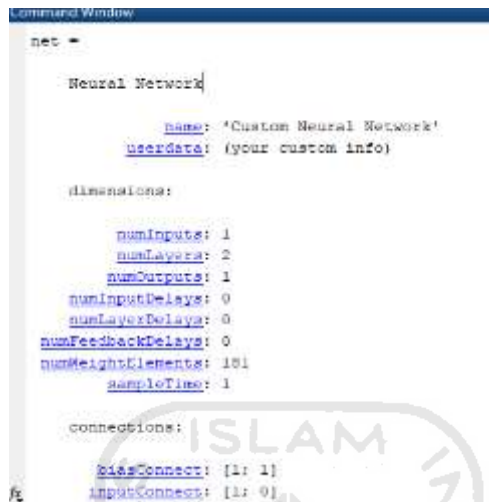
Selanjutnya proses *training* akan berjalan seperti yang terlihat pada Gambar 4. 20.



Gambar 4. 20 Proses *training Backpropagation Neural Network*

Pada Gambar 4. 20 terlihat beberapa informasi proses keseluruhan dengan jumlah 100000 *input* membutuhkan waktu 53 detik dengan *epoch* (pelatihan) 1000, MSE 0,000001

dan batas *gradient* 0,00001 diperoleh *performance* sebesar 0,0780. Sehingga diperoleh diperoleh detail informasi, dimana *best training performance* sebesar 0,077985 pada *epoch* 995. Setelah proses *training* selesai diperoleh hasil keseluruhan dapat diketahui nilai dari *net* yang didalamnya terdapat banyak informasi *output*. Sebagian informasi tersebut dapat dilihat pada Gambar 4. 21.



```

Command Window
net =

Neural Network

    Name: 'Custom Neural Network'
    UserData: (your custom info)

Dimensions:

    numInputs: 1
    numLayers: 2
    numOutputs: 1
    numInputDelays: 0
    numLayerDelays: 0
    numFeedbackDelays: 0
    numWeightsElements: 161
    sampleTime: 1

Connections:

    biasConnect: [1; 1]
    inputConnect: [1; 0]
  
```

Gambar 4. 21 Informasi detail variabel "net"

## 2. Proses pengujian *Backpropagation Neural Network*

Setelah melakukan *training* selanjutnya adalah tahap pengujian dari rekaman suara *known* sebagai contoh rekaman suara pembandingan (P3). Seperti tahap sebelumnya pertama melakukan pengubahan sinyal domain frekuensi dan melakukan *Lowpass filter*. Dimana nilai *training* dengan variabel "net" dilakukan proses pengujian P3 yang dapat dilihat pada Gambar 4. 22

```

%Load Uji File
[y,fs] = audioread('1_P3.wav');
bb = fft(y);
bb2 = abs(bb(1:100000,1))';
%%
% *Proses pengujian*
yin = net(bb2(1,1:100000));
  
```

Gambar 4. 22 Proses pengujian *backpropagation neural network*

Pada Gambar 4. 22 mencari nilai akurasi dari rekaman suara yang dianalisis, proses pencarian dapat dilihat pada Gambar 4. 23.



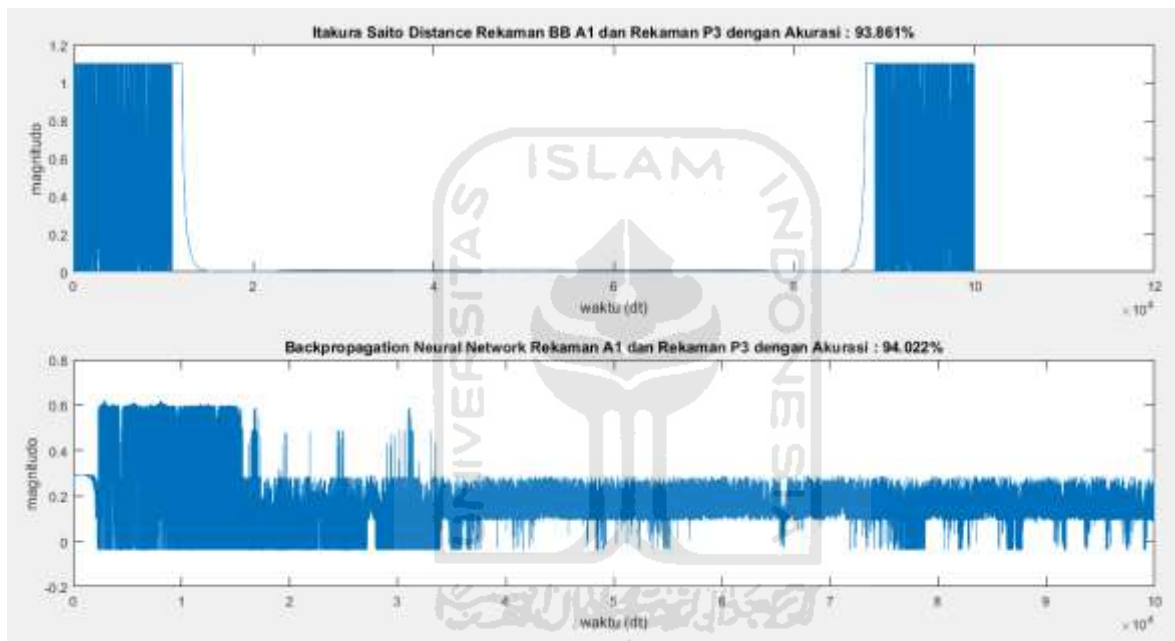
```

%%
% *Hasil*
B = yin(~isnan(yin));
B = B';
lb = abs(b3(1:end-1,1)-B);
disp( [num2str(length(lb(lb<=0.6)))/1000) '%']

```

Gambar 4. 23 Analisis nilai akurasi dari metode *backpropagation neural network*

Berikut hasil dari metode *backpropagation neural network* pada rekaman suara P3 jika ditampilkan dalam sebuah *plot*, sekaligus dibandingkan langsung dengan hasil metode *itakura-saito distance* dapat dilihat pada Gambar 4. 24.



Gambar 4. 24 Hasil Visualisasi metode *itakura-saito distance* dengan metode *backpropagation neural network*

Dilihat dari Gambar 4. 24. rekaman suara A1 dengan rekaman suara P3 dengan metode *Itakura-Saito Distance* sudah dikatakan identik dengan nilai sebesar 93,861%, sedangkan dengan menggunakan metode *Backpropagation Neural Network* mengalami peningkatan akurasi sebesar 94,022%.

#### 4.4. Hasil Analisis Penggunaan Metode Itakura-Saito Distance dengan Metode Backpropagation Neural Network

Berdasarkan hasil implementasi metode *Itakura Saito Distance* dengan metode *Backpropagation Neural Network* dalam mencari nilai akurasi identik suatu rekaman suara

*known* dan rekaman suara *unknown* memperoleh nilai presentase yang dapat menjelaskan suara tersebut identik atau tidak. Berikut hasil analisis dapat dilihat pada Tabel 4. 6.

Tabel 4. 6 Hasil Akurasi identifikasi rekaman suara dengan metode *itakura-saito distance* dan *backpropagation neural network*

No	Subjek Suara		Ket	Presentase (%)	
				Metode <i>Itakura-Saito Distance</i>	Metode <i>Backpropagation Neural Network</i>
				FFT	FFT
				<i>Low-Pass Filter</i>	<i>Low-Pass Filter</i>
				LPC	-
1.	A1	P3	Identik	93,861%	94,022%
2.	A2	P2	Identik	89,312%	90,939%
3.	A3	P4	Identik	95,952%	95,117%
4.	A4	P1	Identik	94,011%	94,432%

Dari hasil keseluruhan yang terdapat pada Tabel 4. 6 rata-rata rekaman suara *unknown* ( $A_n$ ) dan *known* ( $P_n$ ) yang dikatakan identik mengalami peningkatan akurasi. Hal ini sesuai dengan penelitian yang dilakukan sebelumnya oleh (Irawan, 2019) yang membahas tentang perbedaan metode *Itakura Saito Distance* dan metode Konvensional dengan analisa manual statistik (*pitch, formant, spectrogram*) pada 6 rekaman suara perempuan *unknown* dan *known*. Perbedaan metode tersebut dapat dilihat pada Tabel 4. 7, sekaligus pemaparan dari hasil analisa metode *Backpropagation Neural Network* sebagai metode untuk memperkuat hasil akurasi suara dari metode *Itakura Saito Distance* menggunakan 4 rekaman suara laki-laki *unknown* dan *known*.

Tabel 4. 7 Perbedaan metode Konvensional dengan metode *Itakura Saito Distance* dan metode *Backpropagation Neural Network*

Metode Konvensional	Metode <i>Itakura Saito Distance</i> dan Metode BPNN	Variabel
Analisa <i>pitch, formant, spectrogram, likelihood ratio, graphical distribution</i>	Menganalisa nilai <i>formant</i>	Nilai <i>formant 1, 2, .... n</i>
Menganalisa suara dalam bentuk kata	Menganalisa suara dalam bentuk kalimat	Durasi waktu rekaman suara

Tabel 4. 8 Lanjutan

Metode Konvensional	Metode <i>Itakura Saito Distance</i> dan Metode BPNN	Variabel
Lazim/sering digunakan dalam proses <i>audio forensic</i>	Jarang digunakan dalam proses <i>audio forensic</i>	<i>Digital Forensic Analyst Team (DFAT)</i>
Memerlukan <i>tools</i> tambahan dalam menghilangkan <i>noise</i>	<i>Filtering noise</i> secara otomatis	<i>LPC, low-pass filter, dan editing noise audio</i>
Pengucapan kosakata harus jelas (>44,2 Khz, 24 bit)	Mampu menganalisa suara dalam bentuk kalimat dengan kosakata yang kurang jelas (>21 Khz, 16 bit)	Standar kualitas suara berdasarkan bit
Waktu analisa yang dibutuhkan relatif lama	Waktu analisa relatif cepat	<i>Itakura saito dan BPNN</i> durasi rekaman 0.5-2 detik membutuhkan waktu 10-15 menit dengan jumlah <i>frame</i> terbatas. Konvensional durasi rekaman 0.5-2 detik membutuhkan waktu 30-60 menit setiap 1 rekaman suara.
Mampu menampilkan nilai dan grafik disetiap analisa	Memvisualisasikan hasil jarak dari rekaman suara dalam bentuk grafik dan presentase	<i>Itakura saito dan BPNN</i> visualisasi nilai <i>formant</i> dengan ketepatan penentuan parameter sangat mempengaruhi hasil pengujian. Konvensional memvisualisasi nilai, <i>formant, graphical distribution</i> dan <i>spectrogram</i> .

Berdasarkan Tabel 4. 7 berdasarkan parameter dan variable yang ada pada masing-masing metode. Proses analisa metode konvensional secara keseluruhan dilakukan dengan

manual statistik, ada kemungkinan hasil dari tingkat akurasi rekaman suara yang dianalisa kurang maksimal. Sedangkan untuk metode *Itakura Saito Distance* dan BPNN hasil tingkat akurasi rekaman suara maksimal, karena analisa dilakukan secara *automatic* statistik dengan membandingkan jarak nilai *formant*

Perbedaan jenis analisa dan karakter rekaman suara dari metode *Itakura Saito Distance* dan metode Konvensional, sekaligus dibandingkan dengan penggabungan metode *Itakura Saito Distance* dan BPNN dapat dilihat pada Tabel 4. 9.

Tabel 4. 9 Jenis analisa dan karakter dari masing-masing metode

No	Jenis Analisa dan Karakter Rekaman Suara	Metode Konvensional	Metode <i>Itakura Saito Distance</i>	Metode BPNN dan Metode <i>Itakura Saito Distance</i>
1	Kualitas suara baik	✓	✓	✓
2	Kualitas suara yang kurang baik	-	✓	✓
3	Pengucapan kosakata jelas	✓	✓	✓
4	Pengucapan kosakata kurang jelas	-	✓	✓
5	Rekaman suara dengan <i>frame</i> panjang	-	✓	✓
6	Rekaman suara dengan <i>frame</i> pendek	✓	✓	✓
7	Suara yang memiliki <i>noise</i>	✓	✓	✓
8	Pengucapan kosakata yang cepat	-	✓	✓
9	Analisa <i>pitch</i>	✓	-	-
10	Analisa <i>spectogram</i>	✓	-	-
11	Analisa <i>likelihood ratio (LR)</i>	✓	-	-
12	Analisa <i>graphical distribution</i>	✓	-	-
13	Analisa <i>low-pass filter</i>	-	-	✓
14	Analisa FFT	-	-	✓
15	Analisa parameter penurunan <i>gradien</i>	-	-	✓

Berdasarkan Tabel 4. 9 dapat disimpulkan rekaman dengan karakter suara dan kosakata yang baik, pengucapan kosakata yang cepat dan tidak jelas, maupun memiliki *noise* dapat dianalisa menggunakan metode *Itakura Saito Distance* dan BPNN. Sedangkan rekaman suara yang memiliki kualitas baik dan kosakata yang diucapkan dengan jelas, dapat dianalisa dengan metode Konvensional. Berikut hasil akurasi dalam bentuk presentase dari perbandingan 6 rekaman suara perempuan *unknown* dan *known* yang dilakukan penelitian sebelumnya oleh (Irawan, 2019), sekaligus hasil dari penelitian ini dengan 4 rekaman suara laki-laki *unknown* dan *known* dapat dilihat pada Tabel 4. 10.

Tabel 4. 10 Perbandingan hasil akurasi identifikasi rekaman suara *unknown* dan *known*

No	Metode	Analisa pengujian rekaman suara					Bentuk analisa	Jumlah kata
		Pitch	Formant	Likelihood Ratio	Graphical Distribution	Spectrogram		
Metode Konvensional dengan analisa manual statistik								
1	1.1 Analisa Pitch	77,75%					Per-Kata	30 kata
	1.2 Analisa Formant		79,41%				Per-Kata	30 kata
	1.3 Analisa Bandwith		79,41%				Per-Kata	30 kata
	1.4 Analisa Likelihood			79,41%			Per-Kata	30 kata
	1.5 Analisa GD				84,4%		Per-Kata	30 kata
	1.6 Analisa Spectrogram					88,8%	Per-Kata	30 kata
Metode Speech Processing								
2			6/RS	4/RS				
	1.1 LPC		90,83%				Per-kalimat	30 kata
	1.2 Itakura Saito D.		90,83%				Per-kalimat	30 kata
Metode Speech Processing dan Jaringan Syaraf Tiruan								
3			6/RS	4/RS				
	1.1 LPC			95%			Per-kalimat	36 kata
	1.2 Itakura Saito D. dan BPNN			95%			Per-kalimat	36 kata

Note\*:

1. 6/RS adalah 6 rekaman suara perempuan *unknown* dan *known* dari penelitian sebelumnya oleh (Irawan, 2019).
2. 4/RS adalah 4 rekaman suara laki-laki *unknown* dan *known* dari data penelitian.

Nilai presentase pada Tabel 4. 10 diperoleh dari hasil tertinggi dari keseluruhan nilai analisa yang dilakukan penelitian sebelumnya maupun penelitian ini. Dapat disimpulkan bahwa dengan data analisa 6 rekaman suara perempuan *unknown* dan *known* metode Konvensional memiliki nilai presentase 77-88%, sedangkan metode *Itakura Saito Distance* memiliki presentase 90,83%. Untuk penelitian ini dengan 4 rekaman suara laki-laki *unknown* dan *known* menggunakan analisa penggabungan metode *Itakura Saito Distance* dan *Backpropagation Neural Network* presentase sebesar 95%.

Dari hasil keseluruhan perbandingan metode tersebut memiliki kelebihan dan kekurangan masing-masing, sehingga proses analisa perbandingan suara *unknown* dan *known* diharapkan dapat memperkuat hasil akurasi dan *argument* dalam penanganan *audio forensic*. Sehingga metode *Itakura Saito Distance* didukung dengan metode *Backpropagation Neural Network* dapat diterima dalam persidangan yang melibatkan barang bukti rekaman suara.

## BAB 5

### Kesimpulan dan Saran

#### 5.1. Kesimpulan

Berdasarkan hasil yang didapat pada proses implementasi dan analisis metode *Itakura-Saito Distance* dan metode *Backpropagation Neural Network* dapat ditarik beberapa kesimpulan:

1. Metode *itakura-saito distance* sangat dipengaruhi panjang waktu dan kualitas dari rekaman suara tersebut, sehingga memakan waktu yang cukup lama. Sedangkan untuk metode *backpropagation neural network* dengan kondisi yang sama, cukup efektif dalam menganalisis rekaman suara. Kedua metode tersebut memiliki persamaan yaitu untuk mendapatkan hasil yang sesuai harus menggunakan metode lain seperti *fast faurier transform* untuk mengubah sinyal analog ke domain frekuensi dan *low-pass filter* untuk peningkatan kualitas rekaman suara.
2. Mengkombinasikan metode *speech processing* menggunakan *Itakura Saito Distance* dan metode jaringan syaraf tiruan dengan metode *Backpropagation Neural Network* akan mendapatkan nilai akurasi yang lebih efisien atau tinggi dalam mencari rekaman suara yang identik antara suara asli dan suara pembanding, sekaligus memperkuat hasil analisa penanganan *audio forensic*. Hasil nilai akurasi suara yang identik memperoleh hasil nilai presentase tertinggi 95%. Peningkatan tersebut disebabkan dari faktor analisis parameter dari proses *training* yang tepat.

#### 5.2. Saran

Untuk penelitian lebih lanjut mengenai *audio forensic* seperti mengkombinasikan metode *itakura-saito distance* dan metode *backpropagation neural network* untuk menganalisis nilai akurasi pada rekaman suara, diberikan saran diantaranya:

1. Kedua metode tersebut masih sangat tergantung dengan perbaikan kualitas suara dari rekaman suara yang diperoleh, sehingga diperlukan kajian yang lebih lanjut dalam memperbaiki kualitas rekaman suara. Memperbaiki kualitas yang dimaksud seperti penghilangan *noise*, suara yang menggema pada suatu ruangan, ataupun hal lain yang mempengaruhi kualitas suara. Kajian yang dimaksud tanpa melalui alat pengeditan seperti Adobe Audition CC, Praat, dan lain sebagainya.

2. Penelitian ini masih terbatas pada 100000 *frame* yang dianalisis, sehingga masih memungkinkan dikembangkan kembali dengan kajian lebih lanjut dengan mempercepat waktu dalam proses analisisnya.
3. Penentuan nilai parameter pada metode *Backpropagation Neural Network* masih dilakukan secara eksperimen, sehingga hasil analisis tergantung dari ketepatan parameter yang dimasukkan.
4. Penelitian masih terbatas identifikasi rekaman suara laki-laki, sehingga perlu ditambahkan masing-masing suara laki-laki dan suara perempuan dengan sampel data beberapa orang.



## Daftar Pustaka

- Al-Azhar, M. N. (2012). *Digital Forensic (Practical Guidelines for Computer Investigation)*. Jakarta.
- Al-Azhar Nuh, M. (2011). *Audio Forensic : Theory and Analysis*. 1–38.
- Alfina, R., Arifianto, I., Astharini, D., & Wulandari, P. (2019). Mendisain GUI Untuk Menampilkan Nilai FFT dan IFFT Menggunakan LabVIEW. *TESLA: Jurnal Teknik Elektro*, 21(1), 50. <https://doi.org/10.24912/tesla.v21i1.3250>
- Andri, & Tawang, D. A. D. (2018). Validitas Keterangan Ahli dari Penyidik dalam Pembuktian Kasus Ujaran Kebencian Ras DI Media Sosial (Studi Putusan: NO.1105/PID.SUS/2017/PN JKT.UTR). *Jurnal Hukum Adigama*, 1.
- Batenkov, D. (2005). Fast Fourier Transform - Overview. *Key Papers in Computer Science Seminar*. Retrieved from <http://garfield.library.upenn.edu/classics1993/A1993MJ84400001.pdf>
- Bre, F., Gimenez, J. M., & Fachinotti, V. D. (2018). Prediction of wind pressure coefficients on building surfaces using artificial neural networks. *Energy and Buildings*, 158(November), 1429–1441. <https://doi.org/10.1016/j.enbuild.2017.11.045>
- Enqvist, P., & Karlsson, J. (2008). Minimal Itakura-Saito distance and covariance interpolation. *Proceedings of the IEEE Conference on Decision and Control*, 137–142. <https://doi.org/10.1109/CDC.2008.4739312>
- Geeksforgeeks.org. (n.d.). Introduction to Artificial Neural Networks | Set 1. Retrieved May 2, 2020, from <https://www.geeksforgeeks.org/introduction-to-artificial-neural-networks/?ref=lbp>
- Gray, R. M. (2007). Packet speech on the Arpanet: A history of early LPC speech and its accidental impact on the Internet Protocol. *Electrical Engineering*, (June), 1–44.
- Handoko, C. (2017). Kedudukan Alat Bukti Digital Dalam Pembuktian Cybercrime Di Pengadilan. *Jurnal Jurisprudence*, 6(1), 1. <https://doi.org/10.23917/jurisprudence.v6i1.2992>
- Hilmi, M. A. (2014). *Identifikasi Suara Menggunakan Fft Dan Neural Network*. Universitas Muhammadiyah Gresik.
- Huang, H.-X., Li, J.-C., & Xiao, C.-L. (2015). A proposed iteration optimization approach integrating backpropagation neural network with genetic algorithm. *Expert Systems*



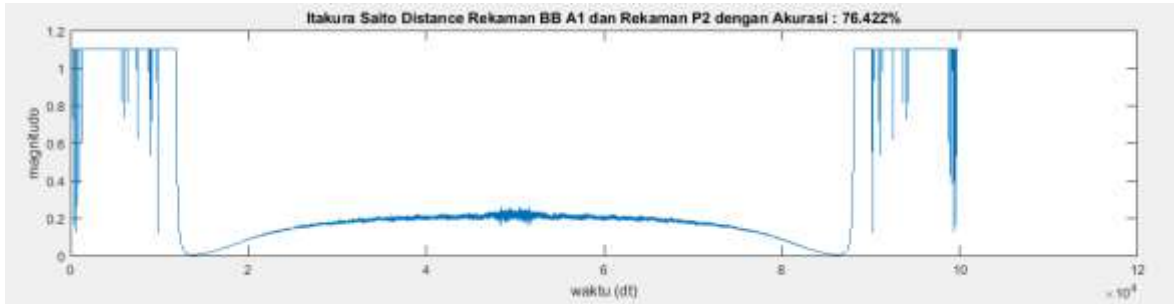
- with Applications*, 42(1), 146–155. <https://doi.org/10.1016/J.ESWA.2014.07.039>
- Hukumonline.com. (2019). Metode “Itakura-Saito”, Silang Pendapat Dua Ahli Forensik di Sidang Tipikor. Retrieved June 13, 2020, from <https://www.hukumonline.com/berita/baca/lt5c5ced64ac1a4/metode-itakura-saito--silang-pendapat-dua-ahli-forensik-di-sidang-tipikor?page=3>
- Imario, A., Sudiharto, D. W., & Ariyanto, E. (2017). Uji Validasi Suara Berbasis Pengenalan Suara (Voice Recognition) Menggunakan Easy VR 3.0. *Prosiding SNATIF Ke-4 Tahun 2017*, 153–160. <https://doi.org/10.2298/PAN0903301G>
- Imran, M., Ali, Z., Bakhsh, S. T., & Akram, S. (2017). Blind Detection of Copy-Move Forgery in Digital Audio Forensics. *IEEE Access*, 5, 12843–12855. <https://doi.org/10.1109/ACCESS.2017.2717842>
- Irawan, A. (2019). *Perbandingan Metode Itakura-Saito Distance dan Manual Statistik (Pitch, Formant, Spectrogram) untuk Akurasi Identifikasi Suara pada Audio Forensik*. Universitas Islam Indonesia.
- Ishii, T., Komiyama, H., Shinozaki, T., Horiuchi, Y., & Kuroiwa, S. (2013). Reverberant Speech Recognition Based on Denoising Autoencoder Division of Information Sciences, Graduate School of Advanced Integration Science, . *Interspeech*, (August), 3512–3516.
- Maher, R. C. (2018). *History of Audio Forensics .Principles of Forensic Audio Analysis*. 29–37. <https://doi.org/10.1007/978-3-319-99453-6>
- Mansyur. (2017). Alat Bukti Rekaman Suara dalam Pembuktian Perkara Tindak Pidana Korupsi. *Jurnal Komunikasi Hukum*, 3(1). <https://doi.org/10.1017/CBO9781107415324.004>
- Mathworks.com. (n.d.). Lowpass Filter Steepness. Retrieved May 2, 2020, from <https://www.mathworks.com/help/signal/ref/lowpass.html>
- Nozaki, Y., & Nakamoto, T. (2017). Itakura-Saito distance based autoencoder for dimensionality reduction of mass spectra. *Chemometrics and Intelligent Laboratory Systems*, 167, 63–68. <https://doi.org/10.1016/J.CHEMOLAB.2017.05.002>
- Padhy, B. K., & Sahu, S. K. (2009). *Analysis of Speech Recognition Techniques*.
- Prasetya, B. W., Susanto, B., & Purwadi, J. (2011). Identifikasi Suara Pria dan Wanita Berdasarkan Frekuensi Suara. *Jurnal Informatika*, 4(1), 2–9. <https://doi.org/10.21460/inf.2008.41.13>
- Prayudi, Y., Luthfi, A., & Pratama, A. M. R. (2014). Pendekatan Model Ontologi Untuk

- Merepresentasikan Body of Knowledge Digital Chain of Custody. *Cybermatika ITB*, 2(2), 36–43. Retrieved from <http://cybermatika.stei.itb.ac.id/ojs/index.php/cybermatika/article/view/67>
- Putri, V. R. C., & Sunarno. (2014). Analisis Rekaman Suara Menggunakan Teknik Audio Forensik Untuk Keperluan Barang Bukti Digital. *Unnes Physics Journal*, 3(1). Retrieved from <https://journal.unnes.ac.id/sju/index.php/upj/article/view/3981>
- Renza, D., Ballesteros L., D. M., & Lemus, C. (2018). Authenticity verification of audio signals based on fragile watermarking for audio forensics. *Expert Systems with Applications*, 91, 211–222. <https://doi.org/10.1016/J.ESWA.2017.09.003>
- Salih, A. O. M. (2017). Audio Noise Reduction Using Low Pass Filters. *OALib*, 04(11), 1–7. <https://doi.org/10.4236/oalib.1103709>
- Sidqi, M. N. (2018). Sistem Cerdas Deteksi Suara Untuk Pengklasifikasian Smart Sound Detection System For Classifying Heart Disease. *Jurnal Elektronik Pendidikan Teknik Elektronika*, 8(1), 42–55.
- Singh, M., & Kumar Garg, E. N. (2014). Audio Noise Reduction Using Butter Worth Filter. *International Journal of Computer & Organization Trends*, 6(1), 20–23. <https://doi.org/10.14445/22492593/ijcot-v6p305>
- Subki, A., Sugiantoro, B., & Prayudi, Y. (2018). Analisis Rekaman Suara Voice Changer dan Rekaman Suara Asli Menggunakan Metode Audio Forensik. *Indonesian Journal on Networking and Security (IJNS)*, 7(1). Retrieved from <http://ijns.org/journal/index.php/ijns/article/view/39/38>
- Sutojo, T., Mulyanto, E., & Suhartono, V. (2011). *Kecerdasan Buatan*. Yogyakarta: Andi Offset.
- Timotius, I. K., Canneva, F., Minakaki, G., Pasluosta, C., Moceri, S., Casadei, N., ... Eskofier, B. (2018). Dynamic footprint based locomotion sway assessment in  $\alpha$ -synucleinopathic mice using Fast Fourier Transform and Low Pass Filter. *Journal of Neuroscience Methods*, 296, 1–11. <https://doi.org/10.1016/J.JNEUMETH.2017.12.004>

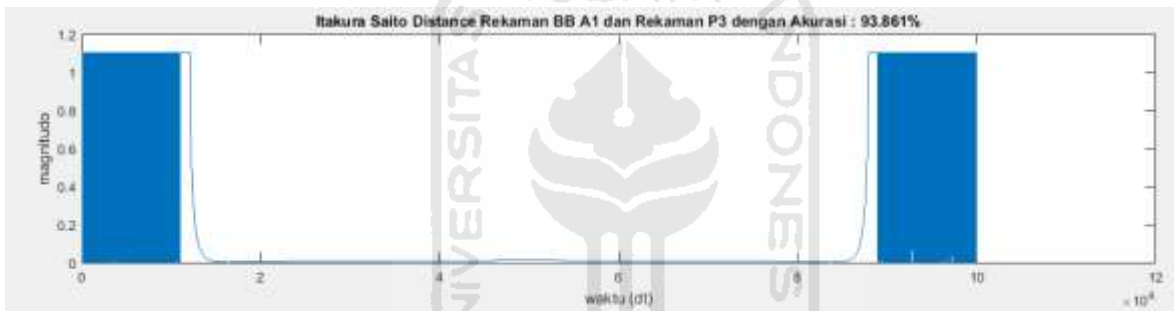
## LAMPIRAN

### Visualisasi Metode Itakura Saito Distance pada 4 Rekaman Suara Laki-Laki Unknown dan Known:

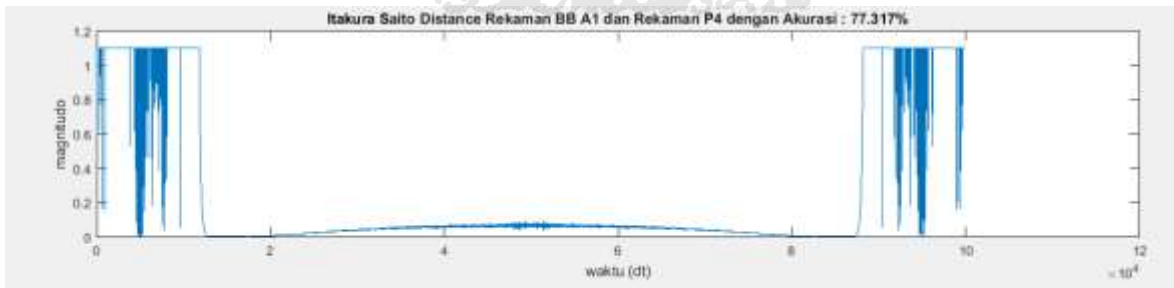
Hasil Visualisasi *Itakura-Saito Distance* Rekaman A1 dan Rekaman P2:



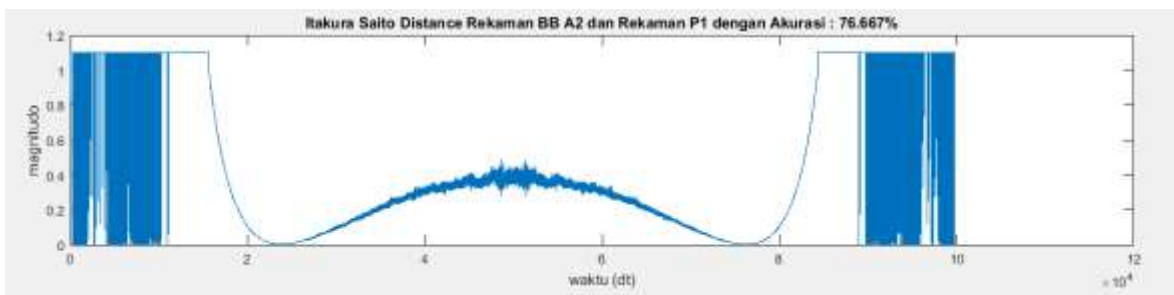
Hasil Visualisasi *Itakura-Saito Distance* Rekaman A1 dan Rekaman P3:



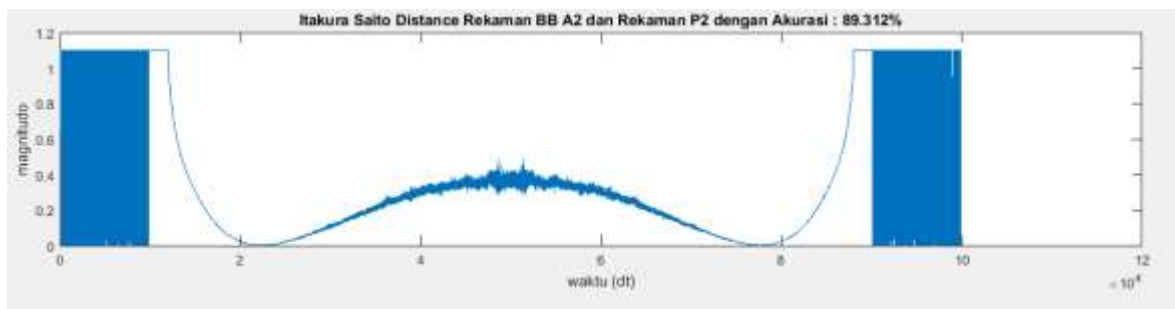
Hasil Visualisasi *Itakura-Saito Distance* Rekaman A1 dan Rekaman P4:



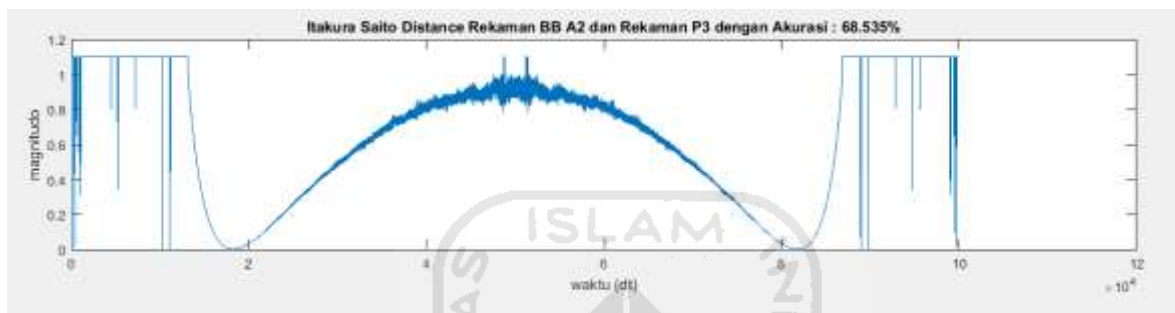
Hasil Visualisasi *Itakura-Saito Distance* Rekaman A2 dan Rekaman P1:



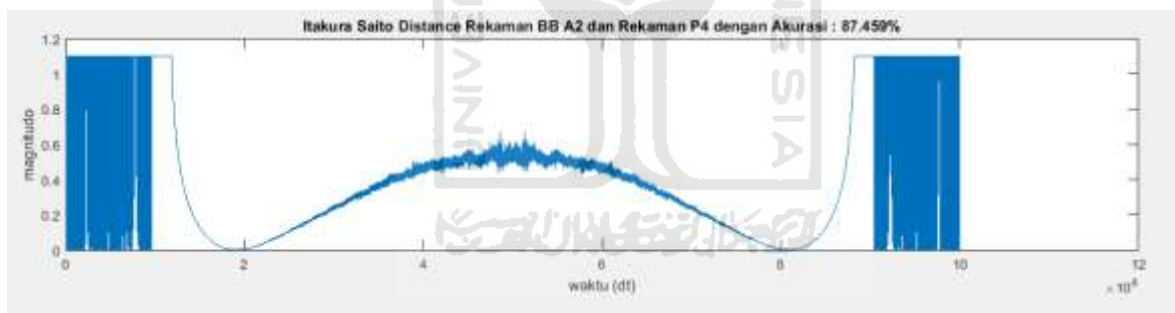
Hasil Visualisasi *Itakura-Saito Distance* Rekaman A2 dan Rekaman P2:



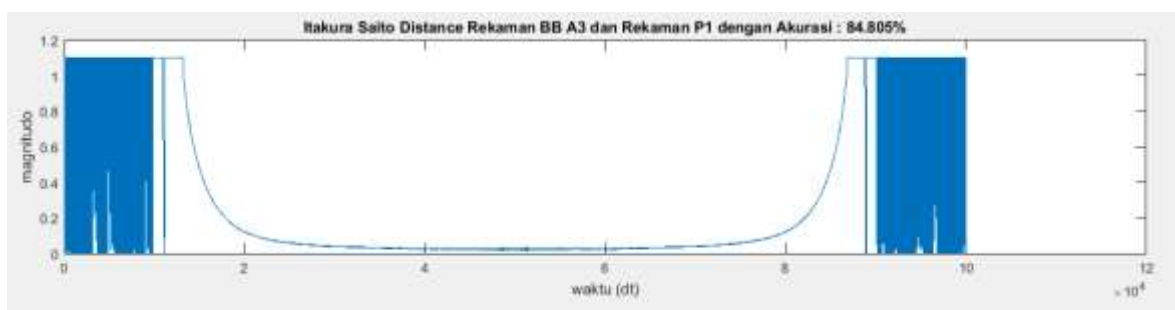
Hasil Visualisasi *Itakura-Saito Distance* Rekaman A2 dan Rekaman P3:



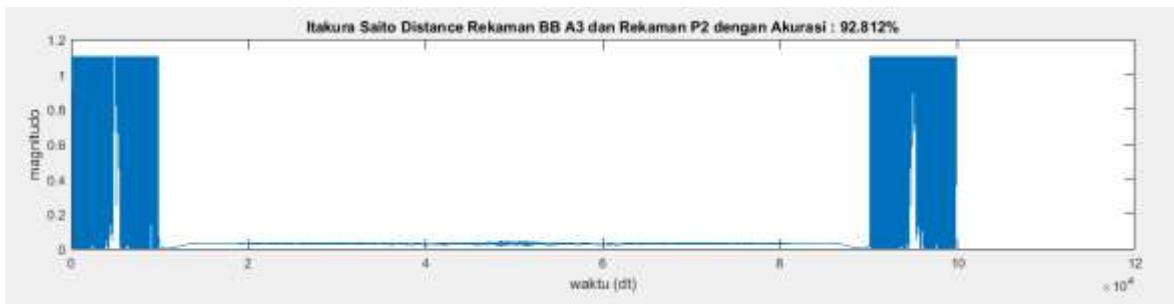
Hasil Visualisasi *Itakura-Saito Distance* Rekaman A2 dan Rekaman P4:



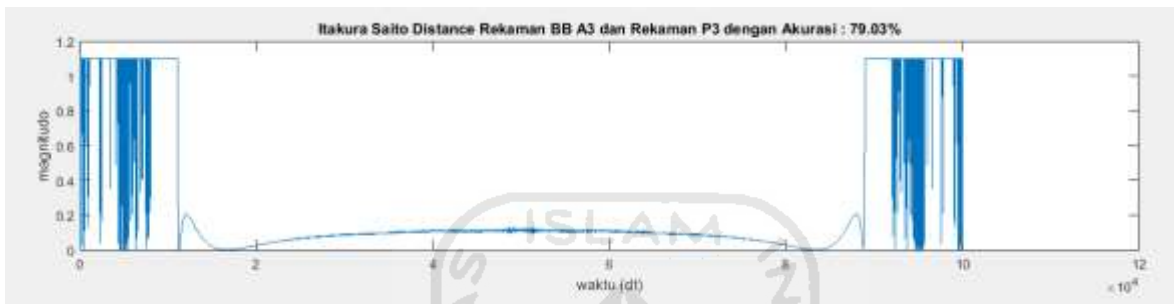
Hasil Visualisasi *Itakura-Saito Distance* Rekaman A3 dan Rekaman P1:



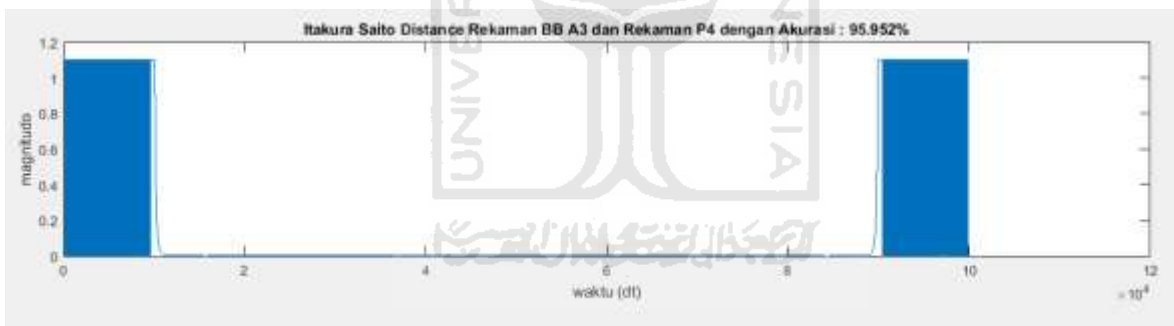
Hasil Visualisasi *Itakura-Saito Distance* Rekaman A3 dan Rekaman P2:



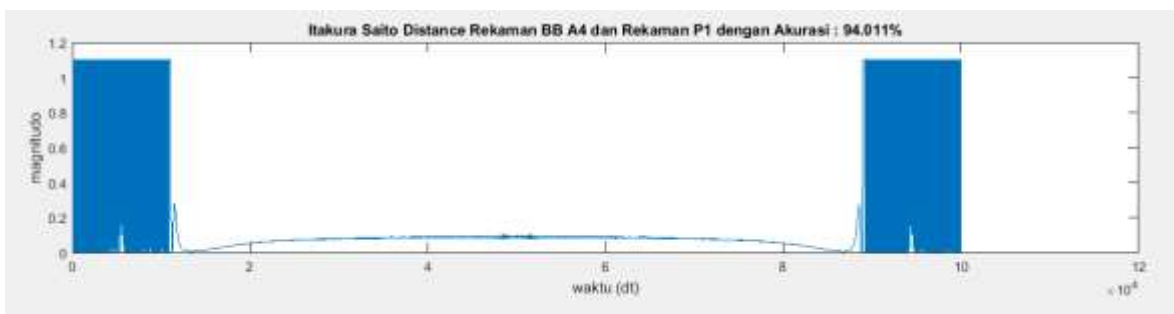
Hasil Visualisasi *Itakura-Saito Distance* Rekaman A3 dan Rekaman P3:



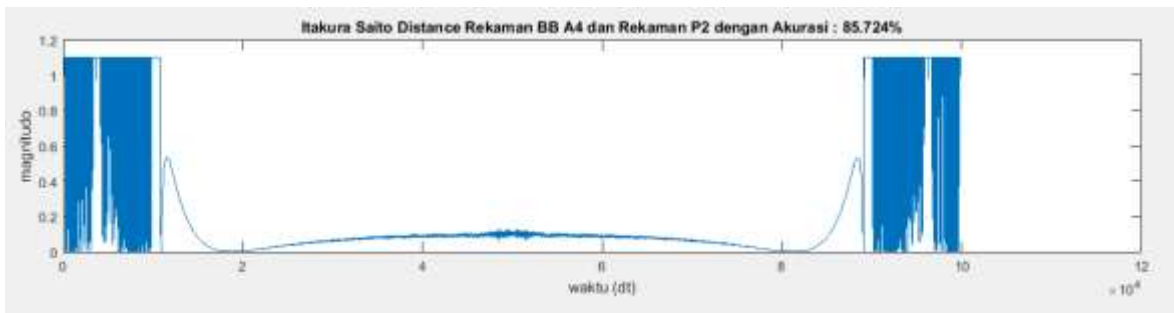
Hasil Visualisasi *Itakura-Saito Distance* Rekaman A3 dan Rekaman P4:



Hasil Visualisasi *Itakura-Saito Distance* Rekaman A4 dan Rekaman P1:



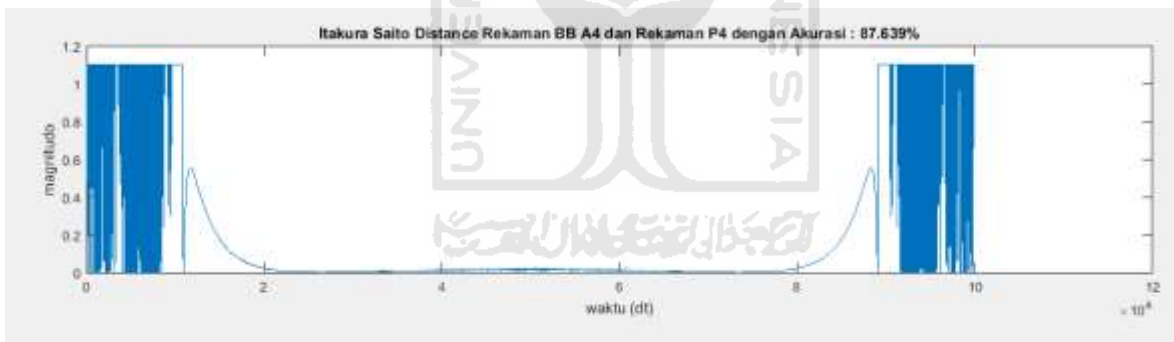
Hasil Visualisasi *Itakura-Saito Distance* Rekaman A4 dan Rekaman P2:



Hasil Visualisasi *Itakura-Saito Distance* Rekaman A4 dan Rekaman P3:

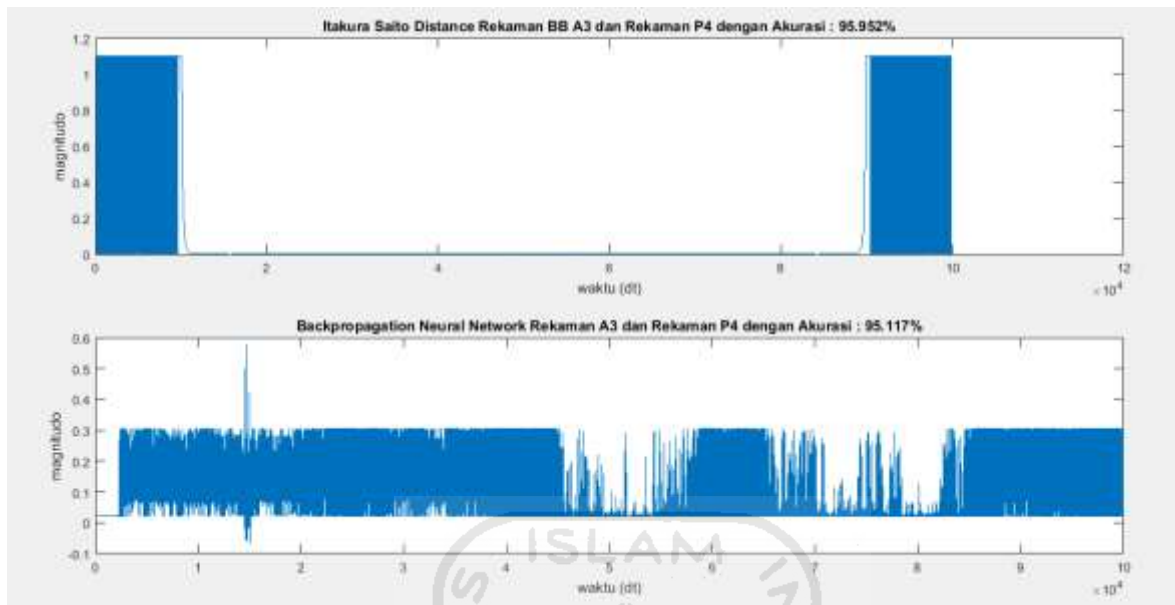


Hasil Visualisasi *Itakura-Saito Distance* Rekaman A4 dan Rekaman P4:



**Hasil Visualisasi metode *itakura-saito distance* dengan metode *backpropagation neural network*:**

Hasil Visualisasi Rekaman A3 dan Rekaman P4:



Hasil Visualisasi Rekaman A4 dan Rekaman P1:

