BAB II
TINJAUAN PUSTAKA

2.1 Landasan Teori

Adapun landasan teori dalam penelitian ini akan dijelaskan secara garis besar dengan teori-teori yang menjadi dasar atau acuan adalah sebagai berikut:

2.1.1 Teori Suara

Menurut Azhar suara yang 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 laring untuk menghasilkan bunyi periodik. Bunyi periodik yang sifatnya konstan tersebut kemudian disaring melalui *vocal tract* (juga disebut dengan isitilah resonator suara atau *articulator*) yang terdiri dari lidah (*tounge*), bibir (*lips*), langit-langit (*palate*) dan lain-lain sehingga bunyi tersebut dapat menjadi keluaran (*output*) berupa bunyi vocal (*vowel*) dan bunyi konsonan (*consonant*) yang membentuk kata-kata yang memeliki arti yang nantinya dapat dianalisis untuk *voice recogntion*.

A. *Spectrum* Suara

Spektrum menggunakan suatu *transform Fourier* cepat (*Fast Fourier Transform, FFT)* matematis untuk melakukan analisis frekuensi. FFT biasanya dinyatakann dengan jumlah point data masukan yang digunakan dalam setiap perhitungan yang selalu berupa kelipatan dua (128, 256, 512, 1024, 2048, atau 4096). Resolusi frekuensi spektrum selalu merupakan nilai cuplikan digital sinyal audio/tutur yang dibagi dengan jumlah point data FFT. Semakin besar jumlah point data FFT, semakin baik resolusi frekuensi spektrum. Frekuensi maksimum yang dihitung oleh FFT dan batas frekuensi tertinggi spektrum adalah setengah nilai cuplikan digital (Rabiner, 1978). Dalam setiap proses analisis spektrum, resolusi waktu dan resolusi frekuensi memiliki hubungan terbalik, resolusi frekuensi yang

sangat bagus berkaitan dengan resolusi waktu yang buruk, sebaliknya resolusi waktu yang sangat bagus berkaitan dengan resolusi frekuensi yang buruk. Hubungan antara resolusi waktu (*Time Resolution*, TR) dalam detik dan resolusi frekuensi (*Frequency Resolution*, FR) dalam Hz adalah:

$$TR = 1/FR \quad (2.1)$$

Penganalisis FFT mentransformasikan data dari domain waktu ke domain frekuensi dengan menghitung FFT. Hal ini didasarkan pada integral Fourier persamaan 2 berikut.

$$y = \int_{-\infty}^{+\infty} x(t)e^{-j2\pi ft}dt \quad (2.2)$$

Dimana :
y = sinyal dalam domain frekuensi (*frequency domain*)
x(t) = sinyal dalam domain waktu (*time domain*)
$e^{-j2\pi ft}$ = adalah konstanta dari nilai sebuah sinyal
f = frekuensi
t = waktu

Namun ini merupakan suatu bentuk yang dapat dihitung secara numeris. Integral ini mensyaratkan bahwa suatu sinyal kontinyu diintegralkan selama waktu yang tak terhingga, tentu saja, diinginkan hasil dalam waktu yang terhingga. Dan karena komputer berhubungan dengan angka, maka diperlukan *digitize* (men-digital-kan) bentuk gelombang, yang dapat membuat waktu bersifat diskrit. Kedua perubahan terhadap sinyal ini mengakibatkan kesalahan dalam spektrum frekuensi yang dihitung. Cuplikan sinyal pada waktu diskrit dapat menyebabkan *aliasing* (yang dapat terlihat sebagai sinyal bayangan (*phantom*) pada tampilan). Pengubahan batas integral panjang tak terhingga menjadi panjang terhingga dapat menyebabkan kesalahan yang disebut kebocoran (yang muncul sebagai energi dari titik tertentu dalam spektrum terbaur (*smeared*) naik dan turun melintasi spektrum). Karena ketidakmungkinan mengukur suatu sinyal untuk waktu yang tak terhingga, maka penganalisis mengubah batas integral ke panjang waktu yang dibutuhkan untuk mengumpulkan blok cuplikan. Blok cuplikan disebut *time record*. FFT mensyaratkan bahwa sinyal dalam *time record* diulang terus-menerus sepanjang waktu. Jika *time record* yang diulang secara aktual tampak seperti sinyal asli, maka tidak akan terjadi kebocoran. Jika, pada sisi lain, tak terlihat seperti sinyal aslinya,

maka terjadi kebocoran. Penerapan fungsi *window* terhadap data yang ada akan dapat membantu mengurangi efek kebocoran dalam domain frekeunsi (Fallside, 1985).

Guna menghitung spektrum y dari sinyal X dapat digunakan FFT dengan menggunakan persamaan berikut:

$$y_{(m)} = \sum_{k=0}^{N-1} X_{(k)}\, e^{2\pi ikm/N} \quad (2.3)$$

Demikian juga untuk menghitung sinyal X dari spektrum y dapat *digunakan Invers Fast Fourier Transformation* (IFFT) dengan menggunakan persamaan berikut:

$$X_{(k)} = \frac{1}{N} \sum_{m=0}^{N-1} y_{(m)}\, e^{-2\pi ikm/N} \quad (2.4)$$

Dengan m= 0……N-1, k = 0….N-1.

Dimana :

$y_{(m)}$ = Spektrum y

$x_{(k)}$ = Sinyal x

$e^{2\pi ikm}$ = konstanta dari nilai sebuah sinyal

2.1.2 Audio Forensik

Audio forensik adalah penerapan ilmu pengetahuna dan metode ilmiah pada barang bukti *digital* audio untuk mendukung upaya penyidikan dan pengungkapan kasus serta membangun fakta-fakta yang diperlukan dalam proses persidangan (Galleh dan Yudi, 2013).

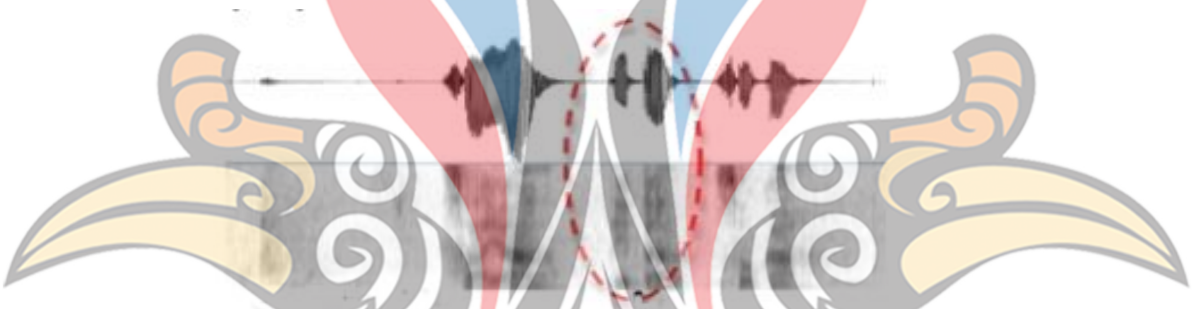
2.1.3 Prosedur Audio Forensik

Menurut Al-Azhar Nuh (2013), untuk mengidentifikasi suara yang berasal dari rekaman barang bukti dan memverifikasinya dengan suara pembanding, berikut adalah tahapan-tahapan yang sesuai dengan *Standard Operational Procedure* (SOP) 12 tentang Analisis Audio Forensik dari *Digital Forensic Analyst Team* (DFAT) pusat laboratorium forensik yang salah satunya mengacu pada *Spectographic Voice Identification*: *A Forensic Survey* yang dikeluarkan oleh *Federal Bureau of Investigation* (FBI), Amerika Serikat:

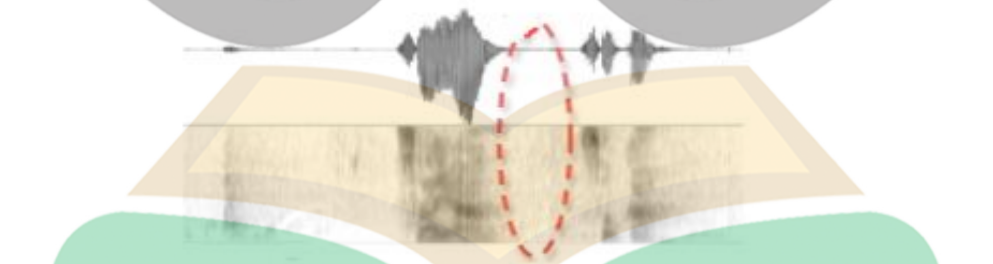
A. *Acquisition*

Acquisition yaitu pengumpulan data barang bukti berupa catatan spesifikasi teknis audio *recorder*, membuktikan keaslian rekaman suara barang bukti berdasarkan fakta yang berkaitan dengan barang bukti, memiliki rekaman suara pembanding yang bebas dari gangguan (*noise*).

Penggunaan rekaman audio sebagai bukti perlu dipastikan keaslian rekaman tersebut, proses ini penting untuk memastikan sedini mungkin bahwa file atau rekaman suara yang digunakan sebagai bukti merupakan rekaman asli dan bukan hasil rekayasa atau hasil modifikasi. Salah satu cara yang digunakan untuk mengecek keaslian rekaman adalah dengan menggunakan ENF (*Electrical Network Frequency*) dan pola perpotongan sinyal menggunakan spektrogram (R.Garg,2013) seperti pada Gambar 2.1 dan Gambar 2.2.



Gambar 2. 1 Rekaman Asli (Sumber : Roy, Dandy, 2016)



Gambar 2. 2 Rekaman Edit (Sumber : Roy, Dandy, 2016)

B. *Audio Enhancement*

Audio Enhancement yaitu perbaikan barang bukti atau mengurangi *noise* pada rekaman barang bukti maka dilakukan proses *enhancement* untuk meningkatkan kualitas rekaman sehingga percakapan terdengar jelas.

C. *Decoding*

Decoding yaitu pembuatan transkrip rekaman yang mencantumkan label subjek dan waktu yang sesuai dengan berjalannya rekaman atau mengekstrasi suara per-kata. Jika suara dalam proses transkrip tidak jelas, maka ditulis ‘tidak jelas’.

D. *Voice recognition*

Voice recognition yaitu proses yang memastikan apakah suara di dalam rekaman barang bukti IDENTIK dengan contoh suara pembanding. Dengan demikian proses ini mengambil kata-kata yang pengucapannya sama antara suara barang bukti denga suara pembanding. Terhadap kata-kata tersebut dilakukan analisis terhadap *pitch, formant* dan *itakura saito distance.*

D.1 Analisis Statistik *Pitch*

Analisis *statistic pitch* dilakukan dengan cara melihat kalkulasi terhadap perbedaan nilai *pitch* dari masing-masing rekaman suara. Untuk menarik kesimpulan dari analisis statistik *pitch* yang mudah dan berargumentasi kuat terlebih dahulu untuk menganalisis nilai *mean* dilanjutkan dengan nilai statistik lainnya apabila jarak perbedaan diantara keduanya >5 Hz maka dinyatakann memiliki perbedaan nilai yang **lebar**. Untuk nilai standar deviasi tidak boleh terlalu dekat dan terlalu tinggi perbedaannya dengan nilai mean.

Tabel 2. 1 Contoh analisis statistik *pitch* kata “percobaan”

Analisis Statistik	Suara percobaan	Suara percobaan_suspect
Pitch Minimum	111.98 Hz	135.67 Hz
Pitch maximum	583.70 Hz	590.35 Hz
Pitch quantile	145.42 Hz	447.88 Hz
Pitch mean	298.89 Hz	380.72 Hz
Pitch standar deviasi	198.05 Hz	154.26 z

D.2 *Analysis Of Variance* (*ANOVA*) *Formant*

Analisis *One-way Anova* (*Analysis of Varience*) yang mengkalkulasi secara statistik nilai-nilai *Formant* 1, *Formant* 2 dan *Formant* 3 dari suara subjek dan *suspect*. Analisis *anova* dapat dihasilkan dengan membedakan dua kelompok data

pada masing-masing *formant* yang ditandai dengan perbandingan *ratio* F dan F *critical*, dan nilai *probability* P dari Hipotesis yang akan diujikan, berikut ini table hasil dari perhitungan analisis *anova*

Tabel 2. 2 Perhitungan analisis *Anova*

Sumber variasi	Jumlah kuadrat	Derajat Bebas	Rerata Kuadrat	F-Uji	F-Tabel
Perlakuan	JKP	(k-1)	RKP=JKP/(k-1)	F=RKP/RKG	$F_{(k-1),(n-k),\alpha}$
Galat	JKG	(n-k)	RKG=JKG/(n-k)		
Total	JKT	(n-1)			

H_0= Tidak ada perbedaan dari kelompok data formant yang di ujikan (*accepted)*

H_1= Ada perbedaan dari kelompok data formant yang diujikan *(rejected)*

Dengan syarat *ratio* F(F-uji) lebih kecil dari F *critical*(F tabel) dan nilai *probability* P (P-*value*) lebih besar dari 0,5. Kemudian dapat disimpulkan bahwa kedua kelompok data yang dianalisis antara suara *subjek* dan *suspect* memiliki tingkat konfidensi 95% serta tidak memiliki perbedaan atau mirip (*accepted*)(Harlan,2018).

Dimana:

JKP : Jumlah kuadrat perlakuan (Treatment sum of squares; SSTr)

JKG : Jumlah kuadrat galat (Error sum of squares; SSE)

JKT : Jumlah kuadrat total (Total sum of squares; SSTo)

RKP : Rerata kuadrat perlakuan (Treatment mean of square; MSTr)

RKG : Rerata kuadrat galat (Error mean of square; MSE)

SAYA
Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Column 1	48	21757.472	453.2806666666667	10392.022416950353
Column 2	42	22082.32	525.7695238095238	7266.530542011614

ANOVA

Source of Variation	SS	df	MS	F	P-value	F critical
Between Groups	117703.81078125711	1	117703.81078125711	13.172122324865079	0.0004763420729797	3.9493210068606906
Within Groups	786352.8058191428	88	8935.827338853896			
Total	904056.6166004	89				

Gambar 2. 3 Contoh tabel *anova* pada *Gnumeric*

Tabel 2. 3 Contoh analisis *anova* kata "percobaan"

Jenis *Formant*	*Ratio F*	*P-Valur*	*F-Critical*	*Conclution*
Formant 1	12.368479	6.3469998E-08	2.6153769	***Rejected***
Formant 2	26.005979	4.197080E-07	3.852409	***Rejected***
Formant 3	28.869103	1.000200E-07	3.852409	***Rejected***
Formant 4	5.716834	1.988796E-07	3.852409	***Rejected***
Formant 5	32.343779	3.407863E-08	3.876788	***Rejected***

D.3 Analisis *Likelihood ratio*

Analisis *Likelihood ratio* bertujuan untuk menguatkan analisis *statistic* nilai *formant* secara mendetail. Analisis *Likelihood ratio* merupakan lanjutan dari analisis *anova formant*. Berikut ini adalah rumus LR sebagai berikut

$$LR = \frac{p(E|Hp)}{p(E|Hd)} \quad (2.5)$$

Dimana:

p (E|Hp): hipotesis tuntutan (*prosecution*), yaitu known dan unknown samples berasal dari orang yang sama

p (E|Hd): hipotesis perlawanan (*defense*), yaitu known dan unknown samples berasal dari orang yang berbeda

p (E|Hp) : p-value *Anova*

p (E|Hd) : 1- *p* (E|Hp)

Pada rumus LR diatas jika LR > 1 maka hal ini mendukung tuntutan p (E|Hp), Sebaliknya jika LR < 1 maka hipotesis perlawanan p (E|Hd) yang didukung. Sehingga dapat disimpulkan bahwa suara *known samples* dan suara *unknown samples* dinyatakann berasal dari orang yang sama harus memenuhi kondisi nilai p (E|Hp) > 0,5.

Berikut ini adalah besarnya *likelihood ratio* yang diikuti dengan verbal statement untuk menjelaskan nilai *likelihood ratio* tersebut, pada Tabel 2.4 dan Tabel 2.5

Tabel 2. 4 Nilai *likelihood ratio* pendukung hipotesis tuntutan p (E | Hp)

LR	*Verbal Statement*	Keterangan
>10,000	*Very strong evidence to support*	Mendukung hipotesis tuntutan *p*(E|Hp)
1000 – 10,000	*Strong evidence to support*	
100 – 1000	*Moderately strong evidenc to support*	
10 – 100	*Moderate evidence to support*	
1 – 10	*Limited evidence to support*	

Tabel 2. 5 Nilai *likelihood ratio* pendukung hipotesis perlawanan p (E| Hd)

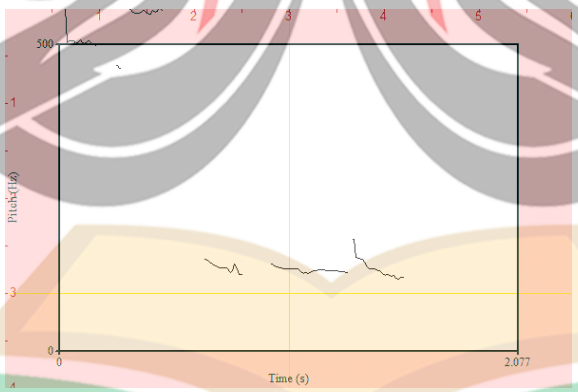
LR	*Verbal Statement*	Keterangan
1-0.1	*Limited evidence to against*	Mendukung hipotesis perlawanan *p*(E|Hp)
0.1-0.01	*Moderate evidence to against*	
0.01-0.001	*Moderately strong evidence to against*	
0.001-0.0001	*Strong evidence against*	
<0.0001	*Very strong evidence against*	

D.4 *Analysis Of Variance* (*ANOVA*) *Itakura Saito Distance*

Analysis Of Variance (ANOVA) Itakura Saito Distance bertujuan untuk mengkalkulasi secara statistik nilai-nilai power *spectral density* dari suara asli dan suara pembanding. *Anova* dapat dihasilkan dengan membedakan dua kelompok data pada masing masing *spectral* suara.

2.1.4 *Pitch*

Menurut Al-Azhar Nuh (2013) *Pitch* adalah frekuensi getar dari pita suara juga disebut dengan istilah frekuensi fundamental (dasar) dengan notasi F_0. Masing-masing orang memiliki *pitch* yang khas (habitual *pitch*) yang sangat dipengaruhi oleh aspek fisiologis laring manusia. 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 (Laver,2002). Frekuensi F_0 ini berubah secara konstan dan memberikan informasi linguistik seseorang seperti pembeda antara intonasi dan emosi. Analisi *pitch* dapat digunakan untuk melakukan *voice recognition* terhadap suara seseorang yang melalui analisis statisik terhadap nilai *Minimum pitch, maximum pitch, mean pitch* dan *standard deviation pitch*. Pada Gambar 2.1 merupakan contoh diagram *pitch* terhadap waktu yang berubah.



Gambar 2. 4 Diagram *pitch* dari kata “percobaan”.

Rumus dasar untuk mencari nilai *pitch* ialah

$$f_0 = \frac{n}{t} \tag{2.1}$$

Dimana

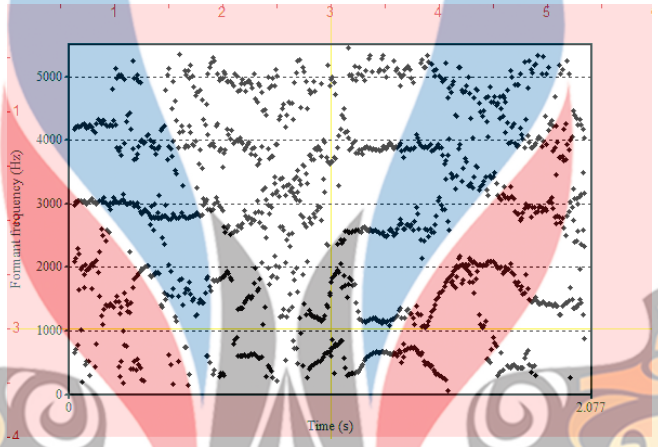
F_0 = Frekuensi dasar (Hz)

n = banyak gelombang

t = waktu

2.1.5 *Formant*

Menurut Al-Azhar Nuh (2013) *Formant* adalah frekuensi-frekuensi resonansi dari filter, yaitu *vocal tract* (*articulator*) yang meneruskaan dan memfilter bunyi keluaran (*output*) berupa kata-kata yang memiliki makna. Secara umum, frekuensi-frekuensi *formant* bersifat tidak terbatas namun, untuk mengidentifikasi seseorang paling tidak ada 3 (tiga) *formant* yang dianalisis yaitu, *Formant* 1 (F_1), *Formant* 2 (F_2) dan *Formant* 3 (F_3). Pada Gambar 2.2 merupakan diagram *formant*.



Gambar 2. 5 Diagram *formant* dari kata "percobaan".

Dengan kata lain *formant* (frekuensi harmonik) adalah frekuensi yang terjadi karena adanya pengaruh (resonansi) dari getaran yang lain sedangkan frekuensi fundamental adalah frekuensi dasar atau frekuensi yang berdiri sendiri dan bukan karena pengaruh resonansi. Rumus dasar untuk mencari *formant* adalah. (Behrman,2018)

$$f_n = (2_n - 1)(c/4L) \tag{2.2}$$

Dimana

$F_{(n)}$ = Nilai *formant* ke-n (Hz)

C = cepat rambat bunyi (m/s), 340 m/s pada suhu ruangan

L = Panjang *vocal tract*, 17.5 cm laki laki dewasa, 15 cm perempuan dewasa, 9 cm anak anak

2.1.6 *Itakura Saito Distance*

Sejak pertama kali di perkenalkan oleh Itakura dan Saito paper yang dipublikasikan pertama kali pada tahun 1968. Metode ini cukup memegang peranan dalam hal kajian mengenai speech coding, analysis, synthesis and recognition. Itakura Saito dikenal juga dengan *The maximum likelihood distortion measure*, pertama digunakan adalah untuk kepentingan analisis spektrum dari sinyal suara. Itakura-Saito pada pinsipnya akan mengukur perbedaan antara spektrum asli dan perkiraan dari spektrum itu. Metode ini secara prinsip akan menggunakan persamaan distribusi *statistic* dari spektrum (dalam Ruang Frekuensi) dari data yang telah dialihkan dari ranah waktu ke ranah frekuensi,

$$D_{is}\left(P_{(\omega),}\acute{P}_{(\omega)}\right) = \int_{-\pi}^{\pi}\left[\frac{P_{(\omega)}}{\acute{P}_{(\omega)}} - \log\frac{P_{(\omega)}}{\acute{P}_{(\omega)}} - 1\right] d_{\omega} \qquad (2.3)$$

Dimana

- $P_{(\omega),}$ Adalah *spectrum* sinyal pertama
- $\acute{P}_{(\omega),}$ Adalah *spectrum* sinyal yang dibandingkan

.

2.2 Penelitian Terdahulu

Tabel 2. 6 Penelitian terdahulu

No.	Nama dan Tahun Publikasi	Hasil
1.	Putri & Sunarno, 2014	Judul : Analisis Rekaman Suara Menggunakan Teknik Audio Forensik Untuk Keperluan Barang Bukti *Digital*. Hasil : • Analisis rekaman suara menggunakan teknik audio forensik dalam penelitian ini dapat menunjukkan secara ilmiah kepemilikan suara pada rekaman, sehingga teknik audio forensik layak untuk digunakan dalam menganalisis rekaman suara untuk menentukan kepemilikan suara sebagai barang bukti *digital*. Namun sampel suara yang dambil hanya 3 sampel dan tidak menunjukan keakuratan metode tersebut.
2.	Aligarh & Bekti, 2016	Judul : Implementasi Metode Forensik dengan Menggunakan *Pitch*, *Formant*, dan Spectrogram untuk Analisis Kemiripan Suara Melalui Perekam Suara Telepon Genggam Pada Lingkungan yang Bervariasi Hasil : • Berdasarkan hasil dari analisis *codec* maka kondisi sepi dan semi-sepilah yang dapat dikatakan mirip dengan pelaku. Sementara lingkungan ramai sulit sekali untuk didapatkankemiripannya, sehingga peluang kejahatan dapat dengan mudah dilakukan pada lingkungan ini. Namun Analisis *codec* dirasa kurang cukup dalam membantu proses identifikasi kemiripan suara tersangka terhadap pelaku
3.	Subki, Sugiantoro, Prayudi, 2018	Judul : Membandingkan Tingkat Kemiripan Rekaman *Voice* Changer Menggunakan Analisis *Pitch*, *Formant* Dan Spectogram. Hasil : • Analisis *pitch*, *formant* dan spectogram dapat digunakan untuk melakukan analisis audio forensik yang terkait dengan rekaman suara *voice* changer. Analisis menggunakan metode *analysis of variance* (*anova*) hanya bisa menganalisis separuh dari keseluruhan suara asli yang sudah dimodifikasi *pitch*-nya

Lanjutan Tabel 2.6 Penelitian Terdahulu

No.	Nama dan Tahun Publikasi	Hasil
4.	Rusydi, Sunardi, Gustafi, 2019	Judul : Analisis Statistik *Pitch* Rekaman Suara Yang Telah Dimanipulasi Menggunakan Audio Forensik Untuk Keperluan Barang Bukti Digital • Analisis statistik *pitch* pada kasus ini dari 20 kata hasil manipulasi suara tersebut tidak identik dengan rekaman suara asli. Terdapat perbedaan nilai analisis statistik *pitch* yang jauh untuk dinyatakan identik dengan suara rekaman asli, nilai yang masih bisa di toleransi sekitar 10 Hz.Analisis statistik pitch tidak efektif dengan kasus manipulasi audio.
5.	Johan Karlsson, Per Enqvist, 2008	Judul *: Minimal Itakura-Saito distance and Covariance interpolation* • Identifikasi *power spectral density* bergantung dari statistik yang diukur seperti, estimasi kovarians. Dalam golongan *power spectral density* konsisten dengan perkiraan spektrum representatif yang dipilih; contoh pilihan tersebut adalah spektrum entropi maksimum dan *Correlogram*. Di sini, kami memilih *power spectral density* sebelumnya untuk mewakili informasi utama, dan *itakura saito distance* dari spektrum terdekat yang dipilih sebelumnya.
6.	Danaba, Arifianto, Rahmadiansyah, 2011	Judul : Pemisahan Sinyal Audio Tercampur dari *Live Music Recording* Multi-SumberMulti-Kanal Dengan Metode *Smooth Itakura Saito NMF* • Perekaman multi sumber music akan mengakibatkan terekamnya sinyal yang tercampur. Sehingga mengakibatkan terjadinya penumpukan *power spectral* maka dilakukan metode *Smooth itakura saito NMF* untuk pemisahan sinyal audio. Setalah dilakukan perhitungan nilai *Mean Square Error* maka perlu dilakukan pengkajian ulang terhadap algoritma dari metode *smooth itakura NMF*.