

Perbandingan Tingkat Kemiripan Rekaman Suara Menggunakan Metode Itakura Saito Distance untuk Mendukung Analisa Audio Forensik

Ardy Wicaksono^{1✉}, Eko Puji Laksono², Selvi Dwi Hartiyani³

^{1,2,3} Prodi Teknologi Informasi, Fakultas Teknik, Universitas Proklamasi 45, Yogyakarta

ardywicaksono@up45.ac.id

Abstract

Audio refers to the sound in the form of electrical or digital signals. Digital audio is often used to record, store, and transmit audio because it can be easily processed or manipulated using computers and other digital devices. Audio is an important medium as one of digital evidence, audio often plays a crucial role in disclosing crime cases so an audio forensics analysis process is needed. The audio contains a recording of a person's voice that has different pronunciation characters, unclear pronunciation, and a lot of noise. There needs to be handling by the Standard Operational Procedure (SOP) for audio forensics. The stages in conducting audio forensics at PUSLABFOR's Digital Forensics Analyst Laboratory Forensic Team (DFAT) consist of 4 stages namely Acquisition, Audio Enhancement, Decoding, and Voice Recognition. This research will discuss audio analysis using the speech processing method, namely the Itakura Saito Distance. This method will determine the level of accuracy of sound and arguments obtained from the analysis process, especially in the handling of audio forensics. Accuracy itself is measured from the value of the closeness of the frequency or spectrum between the original sound recording and the comparison sound recording. Tests were carried out on 4 original voice recordings (unknown) and 4 reference voice recordings (known) using more than 20 vocabularies. The analysis uses a time limit (frame) of 100000 of the total length of time for each vote which shows the highest vote comparison accuracy of 95%.

Keywords: *Itakura Saito Distance, Low-Pass Filter, Linear Prediction, Audio Forensics.*

Abstrak

Audio mengacu pada suara yang berbentuk sinyal listrik atau digital. *Audio* digital sering digunakan untuk merekam, menyimpan, dan mengirimkan audio, karena dapat dengan mudah diproses maupun dimanipulasi menggunakan komputer dan perangkat digital lainnya. *Audio* adalah media penting salah satu barang bukti digital, sering kali bahwa *audio* ini membawa peran krusial untuk mengungkapkan adanya kasus kejahatan sehingga diperlukan proses analisis *audio forensics*. *Audio* tersebut berisi rekaman suara seseorang yang memiliki karakter pengucapan kosakata yang berbeda-beda, pengucapan yang tidak jelas, dan memiliki banyak *noise*. Perlu adanya penanganan yang sesuai dengan *Standart Operational Procedure* (SOP) *audio forensics*. Tahapan dalam melakukan *audio forensics* pada *Digital Forensics Analyst Team* Laboratorium Forensik (DFAT) PUSLABFOR terdiri dari 4 tahapan yakni *Acquisition, Audio Enhancement, Decoding, dan Voice Recognition*. Penelitian ini akan dibahas mengenai analisis *audio* menggunakan metode *speech processing* yaitu Itakura Saito Distance. Metode ini akan mengetahui tingkat akurasi suara dan *argument* yang diperoleh dari proses Analisa, khususnya dalam penanganan *audio forensics*. Akurasi itu sendiri diukur dari nilai kedekatan frekuensi atau *spectrum* antara rekaman suara asli dengan rekaman suara pembandingan. Pengujian dilakukan pada 4 rekaman suara asli (*unknown*) dan 4 rekaman suara pembandingan (*known*) menggunakan lebih dari 20 kosakata. Analisa menggunakan batas panjang waktu (*frame*) sebesar 100000 dari keseluruhan panjang waktu dari setiap suara yang menunjukkan akurasi perbandingan suara tertinggi 95%.

Kata kunci: *Itakura Saito Distance, Low-Pass Filter, Linear Prediction, Audio Forensics.*

KomtekInfo is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.



1. Pendahuluan

Audio merupakan media penting salah satu barang bukti digital, sering kali bahwa *audio* menjadi peran krusial sebagai barang bukti yang diajukan dipersidangan [1]. Secara umum *audio* dihasilkan oleh getaran benda atau suatu benda yang berupa sinyal analog dengan *amplitude* yang berubah secara terus-menerus terhadap waktu dan berkaitan dengan indra pendengaran yang merambat melalui udara [2]. *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* [3]. Suara diakibatkan dari suatu getaran yang menghasilkan pola gelombang tertentu, karena perubahan tekanan udara secara cepat [4]. Suara mempunyai komponen yaitu *pitch* (*level habitual pitch* berkisar pada 50 s/d 250 Hz untuk laki-laki dan 120 s/d 500 Hz untuk perempuan), *formant* (*vocal tract* yang meneruskan dan menfilter bunyi berupa kata-kata), *spectrogram* (representasi *spectral* yang bervariasi terhadap waktu dan membentuk pola umum khusus masing-masing *formant* dalam pengucapan suku kata) yang dapat digunakan

untuk mengidentifikasi karakteristik suara seseorang untuk kepentingan *voice recognition* [5].

Demi memenuhi keuntungan beberapa orang dipengadilan rekaman suara percakapan dari barang bukti *audio* sering sekali dilakukan manipulasi atau pengeditan [6]. 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 forensics* [7]. Kedudukan barang bukti digital sangat penting dalam pembuktian suatu kasus, sehingga diperlukan saksi ahli yang komponen khususnya dibidang keilmuan *audio forensics* [8]. *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 [9]. Seorang saksi ahli harus mampu memaparkan analisis yang relevan menyeluruh dan otentik dalam melakukan serangkaian pemeriksaan. Membuktikan suatu peristiwa yang melibatkan barang bukti *audio* harus sesuai dengan ketentuan, karena menyangkut nasib dari korban maupun tersangka [10]. Penggunaan teknik *audio forensics* layak untuk digunakan dalam menganalisis rekaman suara untuk menentukan kepemilikan suara [11]. Serangkaian pemeriksaan harus mengikuti Undang-Undang, SOP, dan sesuai dengan ketentuan *audio forensics* pada DFAT PUSLABFOR yang mengacu pada *Good Practice Guide for Computer-Based Electronic Evidence* yang diterbitkan oleh *Association of Chief Police Officer (ACPO)* di Inggris, dan *Forensics Examination of Digital Evidence: A Guide for Law Enforcement* di Amerika Serikat [9].

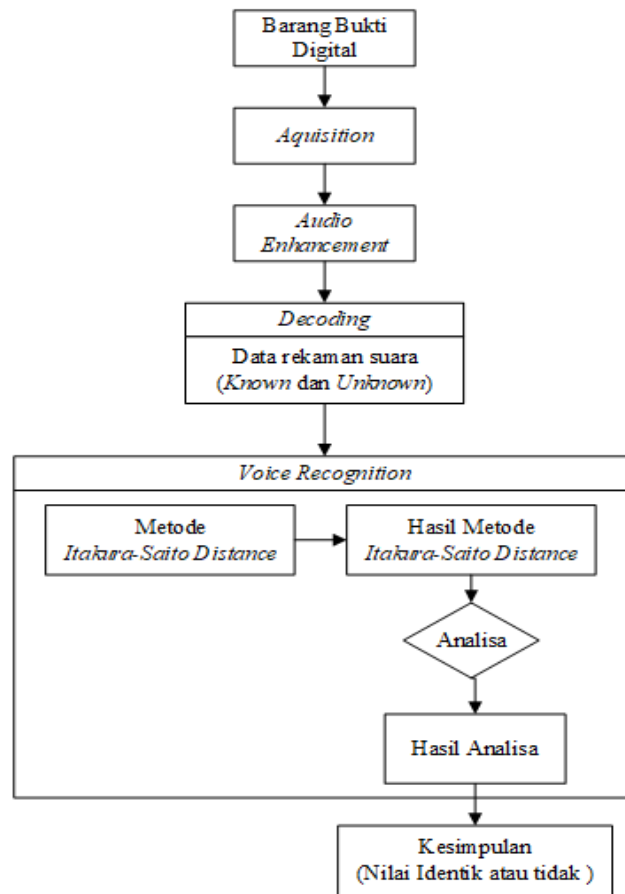
Analisis *audio forensics* tergantung pada ketersediaan rekaman *audio* yang diperoleh. Rekaman *audio* berisi wawancara, penyadapan rahasia, serta merekam interogasi [12]. Rekaman *audio* yang ditemukan memiliki karakteristik berbeda-beda pada setiap orang dan memiliki kualitas belum tentu baik. Rekaman suara dengan merendahkan *pitch* cenderung mudah diidentifikasi, dibandingkan rekaman meninggikan *pitch* [13]. Ketika terjadi tindak pidana dan ditemukan barang bukti berupa rekaman suara terduga tersangka, diperlukan proses analisis untuk menentukan kebenaran rekaman suara yang ditemukan identik. Analisis bukti digital berupa *audio* dapat dilakukan dengan menggunakan metode berbasis *auto encorder* atau *automatic statistic* yaitu *Itakura Saito Distance*. Metode

ini menggunakan jarak antara spektrum suara asli dengan perkiraan suara sebagai indikator akurasi deteksi barang bukti *audio* [14]. *Reverberation autoencorder* merekonstruksi pengenalan ucapan menggunakan spektra jangka pendek dan jangka panjang yang teruji secara matematis untuk meningkatkan akurasi rekaman ucapan yang dihasilkan [15].

Metode *Itakura-Saito Distance* adalah metode *speech processing* yang dapat menggunakan pengkodean *linier prediction* untuk mengukur perbedaan antara spektrum suara asli dan perkiraan spektrum suara. Dari metode perhitungan LPC menggunakan metode *Itakura-Saito Distance*, dapat diperoleh nilai akurasi yang identik apabila semakin kecil jarak antara dua spektrum maka rasio suara semakin mirip [16]. Akurasi sendiri diukur dengan frekuensi atau nilai kedekatan spektral antara rekaman *audio* asli dengan rekaman *audio* pembanding. Persentase akurasi yang tinggi dari metode *Itakura-Saito Distance* masih diperdebatkan. Asumsi tersebut disampaikan oleh seorang ahli kasus korupsi di Indonesia dengan bukti rekaman suara, dimana cara ini banyak digunakan dalam teknologi suara yaitu musik latar atau konser, namun tidak mengarah pada perbandingan dua suara [17]. Metode yang sudah disebutkan diatas diasumsikan mempunyai nilai identik akurasi tinggi dalam membandingkan rekaman suara *unknown* dan *known*. Memperoleh nilai yang signifikan dalam menganalisis bukti digital *file* rekaman suara tentunya akan mempermudah ketentuan barang bukti yang sesuai dengan *Chain of Custody* (CoC).

2. Metodologi Penelitian

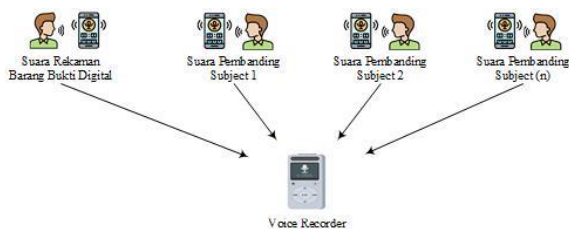
Metode penelitian didasarkan pada tinjauan penelitian sebelumnya pada subjek dengan tujuan untuk mengembangkan lebih lanjut gagasan penelitian yang ada dan menganalisis langkah-langkah apa yang digunakan dalam penggunaan metode yang berlaku dalam pemrosesan bukti *audio forensics*. Penelitian ini untuk menguji coba agar mendapatkan kesimpulan dari pemanfaatan metode apa yang tepat untuk penanganan barang bukti yang berupa rekaman *audio*. Pemahaman mengenai teknik *audio forensics* harus sesuai dengan SOP dari *Digital Forensic Analysis Team* (DFAT), sehingga barang bukti yang diperoleh harus menjamin *Chain of Custody* (CoC). Adapun proses tersebut dapat digambarkan dalam sebuah kerangka penelitian yang dapat dilihat pada Gambar 1.



Gambar 1. Kerangka Penelitian

2.1. Pengambilan Sampel Data

Sesuai dengan SOP tentang analisis *audio forensics*, pengambilan barang bukti dilakukan dengan alat perekam suara dengan pengucapan minimal 20 kata. *Sampling audio* membuat rekaman dengan format *.wav*. Berikut ilustrasi pengambilan sampel rekaman suara dapat dilihat pada Gambar 2.



Gambar 2. Ilustrasi Pemngambilan Sampel

Pengambilan sampel suara pada Gambar 2 digunakan untuk proses implementasi beberapa metode yang akan dikombinasikan dan dianalisis untuk mengetahui tingkat keakuratan suara. Hal tersebut bertujuan untuk melakukan identifikasi rekaman suara dengan mengacu pada rumusan aspek-aspek utama sesuai parameter yang digunakan setiap langkah. Proses analisis menggunakan skema penanganan barang bukti rekaman suara pada barang bukti *audio* dengan tahapan berikut [9]:

1. Acquisition

Proses akuisisi dilakukan dengan menggunakan alat bukti elektronik (*smartphone*) yang meliputi rekaman audio asli yang ditemukan di TKP (*unknown*) serta bukti rekaman yang berfungsi sebagai rekaman audio perbandingan (*known*). Berdasarkan skenario, rekaman audio yang ditemukan pada proses akuisisi *audio recorder* menghasilkan *file image DD* yang ditetapkan, penemuan file yang berekstensi *.wav*.

2. Audio Enhancement

Langkah ini merupakan proses mengevaluasi rekaman *audio* yang ditemukan melalui pemutaran (*playback*) untuk menentukan kualitas suara. Jika kualitas suara rekaman *audio* kurang bagus, mungkin ini penyebabnya faktor kebisingan. Diperlukan suatu proses untuk menghilangkan kebisingan tersebut. Penelitian ini menggunakan algoritma *filter low-pass* menstabilkan *amplifier* dengan menghilangkan frekuensi tinggi dari rekaman audio, sekaligus menggunakan *tools* Adobe Audition CC 2019.

3. Decoding

Langkah ini merupakan proses transkripsi dari rekaman *audio* yang ditemukan dan transkripsi ini dibuat dengan maksud untuk mendapatkan hasil transkripsi yang lebih akurat. Penelitian ini menggunakan proses transkrip online melalui *otransscribe.com*. Transkrip hasil

dimulai dari proses *input* rekaman suara kemudian melakukan transkrip, dan mencari kalimat memiliki lebih dari 20 kata rekaman.

4. Voice Recognition

Langkah ini merupakan teknik yang menganalisis atau mengidentifikasi apakah rekaman *audio* asli (*unknown*) identik dengan rekaman *audio* pembandingan (*known*) yang diperoleh dari tahap *sampling*. Analisis dilakukan dengan Matlab 2015b. Sesuai dengan peraturan *audio forensics*, proses ini membutuhkan setidaknya 20 kata yang menunjukkan kesamaan antara bukti *audio* (*unknown*) dan pembandingan rekaman *audio* (*known*). Penelitian ini akan melakukan analisis pada proses *voice recognition* menggunakan metode *Itakura-Saito Distance*.

2.2. Implementasi Metode

Metode *Itakura-Saito Distance* memiliki beberapa tahapan dalam analisisnya, mulai hasil rekaman suara dari proses *audio enhancement* dengan format *.wav*. Format tersebut akan dijadikan *input* berupa sinyal analog. Sinyal analog tersebut diubah menjadi sinyal frekuensi (*spectrum*) dengan algoritma *Fast Fourier Transform* (FFT)[18]. Parameter pergerakan berbasis FFT termasuk intensitas yang memiliki korelasi yang lebih tinggi [19]. Algoritma FFT merupakan pengembangan dari metode *Discrete Fourier Transform* (DFT), dimana mengubah sinyal analog yang berbasis waktu menjadi berbasis *frequency* [20]. Dimana $y(n)$ untuk *finite-length sequence* yang terdefinisi rentang $0 \leq n \leq N - 1$ dinyatakan pada Persamaan 1 [18].

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

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

Dalam lingkungan yang sangat bising menghilangkan *noise* dapat memulihkan sinyal suara asli, sehingga kosakata akan terdengar jelas. Selanjutnya melakukan langkah *lowpass filter* untuk menghilangkan *high frequency noise* yang terdapat pada proses *audio enhancement* [21]. Untuk menafsirkan *lowpass filter* dapat dilihat pada Persamaan 2.

$$W = (1 - d)x(f_{Nyquist} - f_{pass}) \quad (2)$$

Dimana *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*.

Langkah selanjutnya mengevaluasi kualitas bicara berdasarkan *linear prediction* antara rekaman suara

(*unknown*) diinisialkan tanda A dan rekaman suara pembandingan (*known*) diinisialkan tanda P yang dianalisis dengan melakukan perhitungan LPC.

2.3. Itakura Saito Distance

Metode *Itakura Saito* diusulkan Fumitada Itakura dan Shuzo Saito pada tahun 1960 untuk mengukur kemiripan suara. Metode ini cukup memegang peranan dalam kajian mengenai *speech coding, analysis, synthesis and recognition*. Metode *Itakura Saito* dikenal juga dengan *maximum likelihood distortion distance*, dimana *linear prediction* (LP) dijadikan perkiraan kemungkinan [22].

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 [23]. 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 Persamaan 3 [24].

$$s(n) = a_1s(n - 1) + a_2s(n - 2) + \dots + a_p s(n - p) \quad (3)$$

Dimana koefisien dari parameter (a_1, a_2, \dots, a_p) konstan terhadap panjang *frame* dari suara yang dianalisis. Parameter tersebut untuk mengubah persamaannya dan menambah *variable* $G u(n)$. Model dari *linear prediction coding* (LPC) adalah sampel dari suara asli yang diberikan pada waktu $n, s(n)$, mendekati jarak (p) suara pembandingan seperti terlihat pada Persamaan 4.

$$s(n) = \sum_{i=1}^p a_i s(n - i) + G u(n) \quad (4)$$

Dimana G merupakan *excitation term* yang dimasukkan ke persamaan $G u(n)$. Persamaan $u(n)$ kemudian dinormalisasi *output* dari waktu (n) . Dari sumber eksitasi yang dinormalisasi dipilih nilai karakter ucapan dari suara yang memiliki rangkaian *quasiperiodic*. Selanjutnya menambahkan domain z untuk dijadikan relasi, yang dapat dilihat pada Persamaan 5.

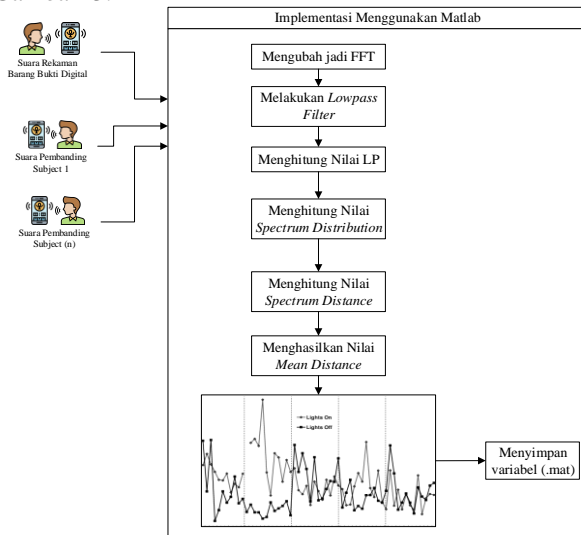
$$H(z) = \frac{S(z)}{G U(z)} = \frac{1}{1 - \sum_{i=1}^p a_i z^{-i}} = \frac{1}{A(z)} \quad (5)$$

Persamaan 5 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 $s(n)$. 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_{15} adalah variabel dari *Itakura Saito* yang dapat dilihat pada Persamaan 6.

$$d_{1S}(S, S') = \int_{-\pi}^{\pi} (e^{V(w)} - V(w) - 1) \frac{dw}{2\pi} \quad (6)$$

$$= \int_{-\pi}^{\pi} \frac{S(w)}{S'(w)} \frac{dw}{2\pi} - \log \frac{\sigma_{\infty}^2}{\sigma_{\infty'}^2} - 1$$

Dimana σ_{∞}^2 spectrum sinyal rekaman suara *unknown* dan $\sigma_{\infty'}^2$ 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. Berikut skema proses implementasi metode *Itakura Saito Distance* dapat dilihat pada Gambar 3.



Gambar 3. Tahap Implementasi metode *Itakura Saito Distance*

Untuk menghitung LPC dibutuhkan rentang waktu tertentu (frame) dan menghitung jarak spectrum antara dua rekaman suara [25]. Selanjutnya akan diperoleh nilai rentan jarak untuk menentukan rekaman suara identik atau tidak, serta terdapat nilai akurasi perbandingannya. Perolehan hasil akurasi tidak lepas dari analisa *anova*. Analisa *anova* merupakan kalkulasi secara statistik nilai-nilai *formant* dari suara *unknown* dan suara *known*. Kalkulasi nilai *formant* menggunakan *Likelihood Ratio* yang dapat dilihat pada Persamaan 7 [26].

$$LR = \frac{p(E|H_p)}{p(E|H_d)} \quad (7)$$

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 [9].

3. Hasil dan Pembahasan

Implementasi metode yang ada diperlukan skenario kasus secara eksperimen dengan melibatkan rekaman suara laki-laki asli (*unknown*) dan rekaman suara laki-laki pembeding (*known*). Ilustrasi kasus dapat berupa cerita yang terjadi sesuai dengan rangkaian ilustrasi saat kejadian perkara yang ditemukan. Melalui tahapan *decoding* diperoleh transkrip isi percakapan rekaman suara yang terduga sebagai tersangka seperti yang dilihat pada Tabel 1.

Tabel 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)		Windows NT

Setelah menentukan kalimat panjang yang akan dijadikan analisis, selanjutnya mencari rekaman suara dari beberapa pembeding. 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.

3.1. Proses *Lowpass Filter*

Rekaman suara yang sudah dilakukan tahap *audio enhancement* akan menjadi *input* berupa sinyal analog. Sinyal analog tersebut diubah kedalam bentuk sinyal frekuensi menggunakan algoritma FFT, kemudian untuk memperbaiki kualitas yang masih terdapat pada langkah *audio enhancement* dilakukan proses *lowpass filter*. Berikut proses untuk melakukan *lowpass filter* pada rekaman suara *a1* dapat dilihat pada Gambar 4.

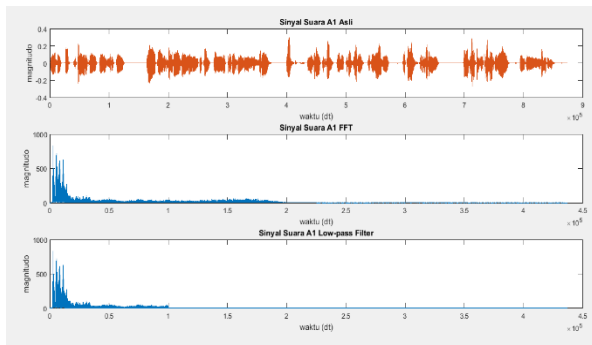
```

3 -   pj = 100000; %%batas
4 -   %%Load Data
5 -   [a1, fs1] = audioread('1_A1.wav');
6 -   %%Proses FFT
7 -   a1 = a1(1:end,1)+a1(1:end,2);
8 -   a1 = a1/2;
9 -   %%Proses low-pass
10 -  a1 = LowPassFilter(a1, 0, pj, '');

```

Gambar 4. Proses melakukan *Lowpass Filter*

Gambar 4 dijelaskan bahwa *input* adalah suara yang dipanggil dengan kode *audioread*, dimana hasilnya dinormalisasi menjadi satu *channel*. 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 Gambar 5.



Gambar 5. Perbaikan Suara dengan Lowpass Filter

Gambar 5 terlihat plot *lowpass filter* yang 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*). Proses selanjutnya ini dinamakan analisis LPC dari parameter waktu atau nilai *frame*.

3.2. Perhitungan Linear Prediction

Pada analisis berbasis LPC, dapat menggunakan *cepstra LPC* untuk menghitung waktu sebagai parameter turunannya. Untuk membuat ekstraksi fitur digunakan nilai absolut yang mengubah sinyal analog menjadi berbasis *frequency*. Kemudian mulai menguji data dari masing-masing rekaman suara tidak diketahui (*unknown*) dan rekaman suara pembanding (*known*). Berikut proses perhitungan LPC dapat dilihat pada Gambar 6.

```
% 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 6. Perhitungan LPC

Gambar 6 menjelaskan bagaimana proses melakukan algoritma FFT dari data *x* yang merupakan algoritma dari *linear prediction*. Kemudian dari algoritma tersebut diperoleh sebuah *function* yang bernama *lpc*. *Function* tersebut akan digunakan untuk mengolah data *audio* dari variabel *a1* dan *p1* hasil dari proses *lowpass filter* yang dapat dilihat pada Gambar 7.

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

y1=y1';
y2=y2';
```

Gambar 7. Perintah Pemanggilan Function LPC

Gambar 7 menjelaskan bagaimana proses *function lpc* berjalan. Pada rekaman suara *a1* dengan membuat variabel baru bernama *y1* dan *p2 = y2*. *Function lpc* akan mengolah variabel data suara dan hasilnya diubah terlebih dahulu menggunakan *function fft* untuk menampilkan hasil berupa sinyal *frequency*. Proses perhitungan *function lpc* tersebut akan mendapatkan nilai rentang panjang waktu yang dapat dilihat pada Gambar 8.

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

y1 =

Columns 1 through 12

    1.0000   -5.8260   13.4221  -13.6653    2.2904    5.6069   -0.3747   -3.5158   -1.2374    1.8739    1.9968   -0.2356

Columns 13 through 24

   -1.7745   -0.9923    0.8570    1.4645    0.1949   -1.2530   -0.9385    0.7665    1.3198   -0.5229   -1.6259    1.7686

Columns 25 through 26

   -0.6997    0.1004
```

Gambar 8. Hasil Nilai LPC y1 dari rekaman A1

Gambar 8 merupakan hasil dari perhitungan *lpc* dengan rentang panjang waktu 100000 (*frame*) dapat disimpan pada sebuah data variabel semisal *a1=y1*, dan seterusnya. Sampel data pada Gambar 8 menampilkan data variabel *y1* jika yang ditampilkan 25 *frame* dari 100000 *frame* yang akan dianalisa. Data pada masing-masing proses *lpc* disimpan, kemudian akan diproses ketahap selanjutnya.

3.3. Proses Itakura Saito Distance

Selanjutnya menentukan jarak *spectrum* antara rekaman suara A1 dengan rekaman suara pembanding P1

menggunakan metode *itakura-saito distance*. Hasil kalkulasi pada Gambar 8 merupakan ekstraksi vektor yang langkah selanjutnya melakukan pengenalan suara menggunakan perhitungan jarak antar vektor. Dimana diproses dengan beberapa *function* seperti *lpcar2ra* digunakan untuk menfilter autokorelasi. *Inverse filter*, *lpcar2rr* untuk mengubah koefisien *autoregressive* menjadi koefisien autokorelasi. *Lpcar2rf* untuk mengonversi koefisien *autoregressive* menjadi refleksi $AR=RF$. Menjadi pokok adalah *function lpcar2ra* yang dapat dilihat pada Gambar 9.

```

[nf,p1]=size(ar);
ra=zeros(nf,p1);
for i=1:p1
    ra(:,i)=sum(ar(:,1:p1+1-i)).
end

```

Gambar 9. Mengkalkulasi LPC dengan *function lpcar2ra*

```

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 10. Menghitung Nilai Jarak Metode *Itakura Saito Distance*

Gambar 10 menjelaskan proses yang didalam *function distisar*. *Function* tersebut berisi beberapa *function* lain seperti *lpcar2ra* sebagai autokorelasi *filtering* dan *lpcar2rr* untuk mengubah koefisien *autoregressive* menjadi koefisien autokorelasi. Kemudian dilanjutkan mengolah *function distisar* yang dapat dilihat pada Gambar 11.

```

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

```

Gambar 11. Proses Menjalankan *Function Itakura Saito Distance*

Gambar 11 merupakan proses mengolah hasil data pada tahap *lpc* sebagai contoh data *y1* dan *y2* yang akan dianalisa menggunakan *function distisar*. Proses ini merupakan proses membandingkan 2 suara untuk

Gambar 9 menjelaskan bagaimana proses *function lpcar2ra*. Pada proses ditunjukkan untuk menfilter autokorelasi *inverse filter*. Tahap ini akan mengolah data suara sebagai contoh data *p1* (suara pembanding) dilakukan *filtering* untuk menghilangkan kemungkinan *noise* masih ada. Selanjutnya *function lpcar2ra* dan *lpcar2rr* akan digunakan untuk mengolah data *p1* ke metode *Itakura Saito Distance* yang dapat dilihat pada Gambar 10.

mencari nilai akurasi. Selanjutnya perintah pencarian suara identik dapat dilihat pada Gambar 12.

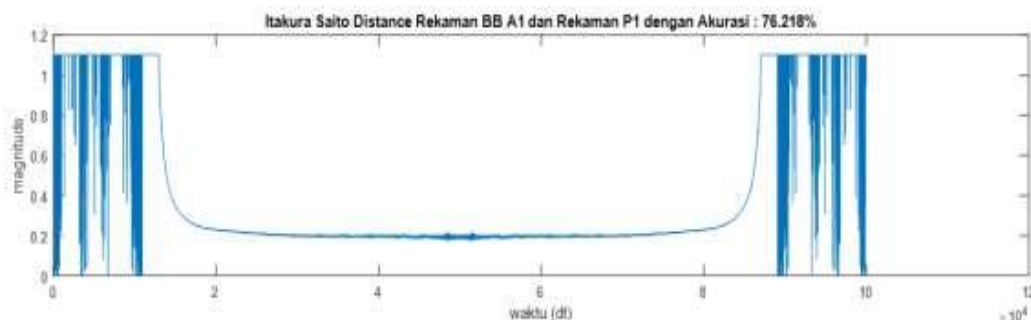
```

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

```

Gambar 12. Perintah Pencarian Akurasi Suara Identik

Gambar 12 merupakan proses mencari hasil nilai *spectrum/formant* dianalisis untuk mencari tingkat akurasi identik atau tidak dari rekaman suara A1 dan rekaman suara pembanding P1. Jika suara yang nilai *formant*-nya mendekati 0 maka suara tersebut identik, dan jika mendekati 1/menjauhi 0 suara tersebut tidak identik. Hasil tersebut dapat dilihat dari visualisasi seperti pada Gambar 13.

Gambar 13. Hasil Visualisasi *Itakura Saito Distance*

Gambar 13 sebagai contoh virtualisasi rekaman suara A1 dan rekaman suara pembanding P1. Virtualisasi tersebut memiliki tingkat akurasi sebesar 76,218% dengan 100000 *frame* dari jumlah keseluruhan *frame*. Proses selanjutnya adalah melakukan analisa perbandingan suara satu-persatu antara rekaman suara asli (*known*) dan rekaman suara pembanding (*unknown*).

3.4. Hasil Analisis Metode *Itakura Saito Distance*

Berdasarkan proses satu-persatu yang sudah dilakukan dalam mencari nilai akurasi identik dari suatu rekaman suara *known* dan rekaman suara *unknown* memperoleh nilai presentase yang dapat menjelaskan suara tersebut identik atau tidak. Berikut hasil dari empat rekaman suara asli (*unknown*) dari empat responden dan empat rekaman suara pembanding (*known*) dari empat responden, dapat dilihat pada Tabel 2.

Tabel 2. Hasil Evaluasi Perbandingan Nilai *Itakura Saito Distance* pada rekaman *unknown* (A_n) dan rekaman *known* (P_n)

Subjek Rekaman Suara	Presentase Akurasi	Keterangan
A1	P1	76,218%
	P2	76,423%
	P3	93,861%
	P4	77,317%
A2	P1	76,667%
	P2	89,312%
	P3	68,535%
	P4	87,459%
A3	P1	84,805%
	P2	92,812%
	P3	73,030%
	P4	95,952%
A4	P1	94,011%
	P2	85,724%
	P3	83,763%
	P4	87,639%

Dilihat pada Tabel 2 rata-rata rekaman suara rekaman suara *unknown* (A_n) dan *known* (P_n) yang dikatakan identik mengalami akurasi suara perbandingan sebesar 95%. Proses analisa perbandingan suara unknown dan known diharapkan dapat memperkuat hasil akurasi dan argument dalam penanganan *audio forensics*. Sehingga metode *Itakura Saito Distance* dapat diterima dalam persidangan yang melibatkan barang bukti rekaman suara.

4. Kesimpulan

Metode *Itakura-Saito Distance* sangat dipengaruhi panjang waktu dan kualitas dari rekaman suara tersebut, sehingga memakan waktu yang cukup lama dalam proses analisisnya. Untuk mendapatkan hasil yang sesuai harus menggunakan metode lain seperti *Fast Fourier Transform* untuk mengubah sinyal analog ke domain frekuensi dan *Lowpass Filter* untuk peningkatan kualitas rekaman suara. Implementasi metode *speech processing* menggunakan *Itakura Saito Distance* dengan mengikuti tahapan penanganan barang bukti *audio forensics* berdasarkan skenario kasus sesuai SOP seperti tahapan *acquisition*, *audio enhancement*, *decoding*, dan *voice recognition*, memiliki hasil nilai akurasi suara dari rekaman suara asli (*unknown*)/suara tidak diketahui dan rekaman suara pembanding (*known*)/suara yang diketahui memperoleh presentasi tertinggi 95%. Nilai akurasi yang tinggi tersebut diharapkan dapat mencari rekaman suara yang identik antara suara yang tidak diketahui dengan suara pembanding, sekaligus memperkuat *argument* dalam penanganan barang bukti *audio* khususnya rekaman suara manusia.

Daftar Rujukan

- [1] T. Feri Efendi, "The Management of Physical Evidence and Chain of Custody (CoC) in Digital Forensic Laboratory Storage," *International Journal of Seecology*, vol. 1, no. 1, 2019.
- [2] A. Imario, D. W. Sudiharto, and E. Ariyanto, "Uji Validasi Suara Berbasis Pengenalan Suara (Voice Recognition) Menggunakan Easy VR 3.0," *Prosiding SNATIF ke-4 Tahun 2017*, pp. 153–160, 2017, doi: 10.2298/PAN0903301G.
- [3] Mansyur, "Alat Bukti Rekaman Suara dalam Pembuktian Perkara Tindak Pidana Korupsi," *Jurnal Komunikasi Hukum*, vol. 3, no. 1, 2017, doi: 10.1017/CBO9781107415324.004.
- [4] A. Sri Irtawaty, M. Ulfah, and E. Reina Elfira Tamzil, "Peningkatan Performansi aplikasi pembeda Suara Laki-Laki Dan Perempuan Berdasarkan Frekuensi Dengan Menggunakan metode Fuzzy Logic," *Jurnal ECOTIPE*, vol. 7, no. 1, pp. 1–6, 2020, doi: 10.33019/ecotipe.
- [5] A. Subki, B. Sugiantoro, and Y. Prayudi, "Membandingkan Tingkat Kemiripan Rekaman Voice Changer Menggunakan Analisis Pitch, Formant dan Spectrogram," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 1, p. 17, Feb. 2018, doi: 10.25126/jtiik.201851500.
- [6] M. Imran, Z. Ali, S. T. Bakhsh, and S. Akram, "Blind Detection of Copy-Move Forgery in Digital Audio Forensics," *IEEE Access*, vol. 5, pp. 12843–12855, 2017, doi: 10.1109/ACCESS.2017.2717842.
- [7] D. Renza, D. M. Ballesteros L., and C. Lemus, "Authenticity verification of audio signals based on fragile watermarking for audio forensics," *Expert Syst Appl*, vol. 91, pp. 211–222, 2018, doi: 10.1016/j.eswa.2017.09.003.
- [8] C. Handoko, "Kedudukan Alat Bukti Digital Dalam Pembuktian Cybercrime Di Pengadilan," *Jurnal Jurisprudence*, vol. 6, no. 1, p. 1, 2017, doi: 10.23917/jurisprudence.v6i1.2992.
- [9] M. Al-Azhar Nuh, "Audio Forensic: Theory and Analysis," pp. 1–38, 2011.
- [10] Andri and D. A. D. Tawang, "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*, vol. 1, 2018.
- [11] H. S. Pratama Nawawi and D. Yusup, "Analisis Penemuan Barang Bukti Digital Melalui Rekaman Suara Menggunakan Praat Dengan Metode Audio Forensik," vol. 7, no. 1, 2022, doi: <https://doi.org/10.32528/justindo.v7i1.5368>.
- [12] R. C. Maher, "History of Audio Forensics. Principles of Forensic Audio Analysis," pp. 29–37, 2018, doi: 10.1007/978-3-319-99453-6.
- [13] A. Subki, B. Sugiantoro, and Y. Prayudi, "Analisis Rekaman Suara Voice Changer dan Rekaman Suara Asli Menggunakan Metode Audio Forensik," *Indonesian Journal on Networking and Security (IJNS)*, vol. 7, no. 1, 2018, [Online]. Available: <http://ijns.org/journal/index.php/ijns/article/view/39/38>
- [14] Y. Nozaki and T. Nakamoto, "Itakura-Saito distance based autoencoder for dimensionality reduction of mass spectra," *Chemometrics and Intelligent Laboratory Systems*, vol. 167, pp. 63–68, Aug. 2017, doi: 10.1016/J.CHEMOLAB.2017.05.002.
- [15] T. Ishii, H. Komiyama, T. Shinozaki, Y. Horiuchi, and S. Kuroiwa, "Reverberant Speech Recognition Based on Denoising Autoencoder Division of Information Sciences, Graduate School of Advanced Integration Science," *Interspeech*, no. August, pp. 3512–3516, 2013.
- [16] A. Irawan, "Perbandingan Metode Itakura-Saito Distance dan Manual Statistik (Pitch, Formant,

- Spectrogram) untuk Akurasi Identifikasi Suara pada Audio Forensik,” Universitas Islam Indonesia, 2019.
- [17] Hukumonline.com, “Metode ‘Itakura-Saito’, Silang Pendapat Dua Ahli Forensik di Sidang Tipikor”, 2019. <https://www.hukumonline.com/berita/baca/lt5c5ced64ac1a4/metode-itakura-saito--silang-pendapat-dua-ahli-forensik-di-sidang-tipikor?page=3> (accessed Jun. 13, 2020).
- [18] H. C. Lin and Y. C. Ye, “Reviews of bearing vibration measurement using fast Fourier transform and enhanced fast Fourier transform algorithms,” *Advances in Mechanical Engineering*, vol. 11, no. 1, Jan. 2019, doi: 10.1177/1687814018816751.
- [19] I. K. Timotius *et al.*, “Dynamic footprint-based locomotion sway assessment in α -synucleinopathic mice using Fast Fourier Transform and Low Pass Filter,” *J Neurosci Methods*, vol. 296, pp. 1–11, Feb. 2018, doi: 10.1016/J.JNEUMETH.2017.12.004.
- [20] R. Alfina, I. Arifianto, D. Astharini, and P. Wulandari, “Mendisain GUI Untuk Menampilkan Nilai FFT dan IFFT Menggunakan LabVIEW,” *TESLA: Jurnal Teknik Elektro*, vol. 21, no. 1, p. 50, 2019, doi: 10.24912/tesla.v21i1.3250.
- [21] A. O. M. Salih, “Audio Noise Reduction Using Low Pass Filters,” *Oalib*, vol. 04, no. 11, pp. 1–7, 2017, doi: 10.4236/oalib.1103709.
- [22] R. M. Gray, “Packet speech on the Arpanet: A history of early LPC speech and its accidental impact on the Internet Protocol,” *Electrical Engineering*, no. June, pp. 1–44, 2007.
- [23] D. Bolin and K. Kirchner, “Equivalence of measures and asymptotically optimal linear prediction for Gaussian random fields with fractional-order covariance operators,” *Bernoulli*, vol. 29, no. 2, May 2023, doi: 10.3150/22-BEJ1507.
- [24] B. K. Padhy and S. K. Sahu, “Analysis of Speech Recognition Techniques,” BTech thesis, 2009.
- [25] V. Kanabur, S. S. Harakannanavar, and D. Torse, “An Extensive Review of Feature Extraction Techniques, Challenges and Trends in Automatic Speech Recognition,” *International Journal of Image, Graphics and Signal Processing*, vol. 11, no. 5, pp. 1–12, May 2019, doi: 10.5815/ijgisp.2019.05.01.
- [26] P. N. Robinson *et al.*, “Interpretable Clinical Genomics with a Likelihood Ratio Paradigm,” *Am J Hum Genet*, vol. 107, no. 3, pp. 403–417, Sep. 2020, doi: 10.1016/j.ajhg.2020.06.021.