CyberSecurity and Digital Forensics e-ISSN: 2615-8442
Vol. 5, No. 2, November 2022, p.66-70

PITCH STATISTICAL ANALYSIS OF FORENSIC AUDIO FOR DIGITAL EVIDENCE

Rahmawati Nafi'ah

Master of Informatics, Faculty of Science and Technology
UIN Sunan Kalijaga, Yogyakarta
rahmawati.nafiah@gmail.com

Abstract

Information technology and computers continue to evolve and have been widely and deeply utilized. Computer / digital forensics are applications from computer technology for legal evidence related to high-tech crimes to be able to obtain digital evidence to ensnare criminals. The area of digital forensics is quite extensive because there are many types of digital content, such as websites, e-mails, images, audio, video and others. In both criminal and civil cases evidence is required, one of which is digital / electronic evidence. Evidence in the form of an audio recorder which produces a voice recording of a person's conversation with another is one of the electronic evidence. The identity of recorded sound through audio forensic examination can be identified by means of a comparison method. This method compares the sound in the recorded evidence with the sound recorded for comparison. If the results are identical, it can be concluded that the recorded evidence comes from the owner of the comparator vote. Pitch analysis is based on statistical calculation of the Pitch value of each of the unknown (Original Voices) and known (Original Voices with Voice changer) voices. Based on experiments that have been carried out on examples of 20 spoken words, it is found that there is a difference in the value of the Pitch statistical analysis of the original voice recording with the voice given by the voice changer. Pitch analysis is based on statistical calculation of the Pitch value of each of the unknown (Original Voices) and known (Original Voices with Voice changer) voices. Based on experiments that have been carried out on examples of 20 spoken words, it is found that there is a difference in the value of the Pitch statistical analysis of the original voice recording with the voice given by the voice changer. Pitch analysis is based on statistical calculation of the Pitch value of each of the unknown (Original Voices) and known (Original Voices with Voice changer) voices. Based on experiments that have been carried out on examples of 20 spoken words, it is found that there is a difference in the value of the Pitch statistical analysis of the original voice recording with the voice given by the voice changer.

Keywords: Pitch analysis, audio forensics, digital evidence.

1. INTRODUCTION

Information and computer technology continues to develop and has been utilized extensively and deeply. Computer/digital forensics is the application of computer technology to prove laws related to high-tech crimes so that digital evidence can be obtained to ensnare the perpetrators of crimes.

According to(Nigroho & Al-Azhar, 2017)Digital forensics is the use of analytical and investigative techniques to identify, collect, examine and store evidence/information that is magnetically stored/encoded on a computer or digital storage media as evidence in uncovering criminal cases that can be legally justified.

The area of digital forensics is quite broad because there are many types of digital content, such as websites, emails, images, audio, video, and others. In both criminal and civil cases, evidence is required, one of which is digital/electronic evidence. Evidence in the form of an audio recorder (sound recording device) that produces voice recordings of one person's conversations with another person is one of the electronic evidences. As the 2016 ITE Law article 1 states that sound recording is one of the valid digital evidence used to resolve a criminal case(Subki et al., 2018) (Rusydi et al., 2019).

Sound recording evidence through audio forensic examination can be identified by the comparative method. This method compares the sound in the recorded evidence with the recorded sound as a comparison. If the results are identical, it can be concluded that the recorded evidence came from the owner of the comparator's voice.

In this study discussed Pitch statistical analysis. This research library is based on a literature review of research texts related to audio forensics.

2. LITERATURE REVIEW

2.A. Forensic Audio

Audo or sound is a way of human communication. The human voice is something unique and has the characteristics of the owner of the voice. Progress Current technology allows human

voices to be used as commands on computer devices (Umar et al., 2019).

Audio forensics is a field of science that analyzes audio such as sound recordings. Sound recordings always have information in the form of frequency characteristics, the identity of these frequencies can be identified(Huizen et al., 2017) (Rusydi et al., 2019).

2.B. Sound Components

Sound is produced through the process of Generation and Filtering. Generation process, sound is first produced by vibrating the vocal cords in the larynx (larynx) to produce periodic sounds. The constant periodic sound is then filtered through the vocal tract which consists of the tongue, teeth, lips, palate and others so that the sound can become an output sound in the form of a vowel sound and or a consonant sound.(Al Azhar, 2012).

Voice consists of several components, namely Pitch, formant and spectrogram which can be used to identify the characteristics of a person's voice for voice recognition purposes.

1. *pitches*

*pitches*is the high and low pitch in a sound. In the human voice, Pitch is generated by a vibrational frequency called the fundamental frequency which has the notation 0(Deva & Mardianto, 2019). Vibration frequency of the vocal cords / fundamental (basic) frequency with the notation F0. Each person has a distinctive Pitch (habitual Pitch) which is strongly influenced by the physiological aspects of the human larynx(Al Azhar, 2012).

2. Formant

Formants are the resonant frequencies of the filter, namely the vocal track which forwards and filters the periodic sound of the vocal cords vibrating into output sounds in the form of words that have meaning.(Al Azhar, 2012). Formant is the resonant frequency of the filter, namely the vocal tract (articulator) which transmits and filters periodic sounds from the vibration of the vocal cords into the output sound of words. Formant is the highest frequency energy in sound(Mehmet Mehdi Karakoc, 2017)(Deva & Mardianto, 2019).

In general, formant frequencies are unlimited, generally drawn from Formant 1 (F1) to Formant 5 (F5), but for the identification of a person's voice, at least 3 (three) formants are analyzed, namely Formant 1 (F1).), Form 2 (F2) and Form 3 (F3).

3. Spectrogram

The spectrogram is a time-varying spectral representation that shows the level of spectral density (energy intensity). This energy level is known as the formant bandwidth. Spectrogram by some experts is also known as voice fingerprint.(Al Azhar, 2012). Spectograms are also known as voice fingerprints because they have detailed things in them(Deva & Mardianto, 2019).

2.C. Audio Forensic Procedures

The following steps are used in accordance with the Standard Operating Procedure (SOP) 12 concerning Audio Forensic Analysis from the Digital Forensic Analyst Team (DFAT) Puslabfor which refers to one of the Spectrographic Voice Identification: A Forensic Survey issued by the FBI (Federal Bureau of Investigation), United States of America(Nigroho & Al-Azhar, 2017).

1. *Acquisition*

Collection of audio recorders to be analyzed, comparative sounds (control or unknown samples) and complete administrative investigations.

2. *Audio Enhancement*

If the audio quality is not good, an enhancement process will be carried out to improve the quality of the recording so that the speech in the voice recording can be heard clearly.

3. Decoding

Making transcripts of recordings by at least 2 (two) examiners to get a more precise accuracy value of the transcript results.

4. *Voice recognition*

This process ensures whether the voices in the recorded evidence are IDENTICAL with the comparative voice samples. This voice recognition must obtain a minimum of 20 (twenty) words that have different meanings and can be accepted because they have very similar patterns and analysis to conclude that the voice of the evidence is IDENTICAL with the voice of the comparator.

If these words are not found, then the conclusion is NOT IDENTICAL. If the number of words in the recorded evidence does not reach 20 (twenty) words with different meanings that can be analyzed, then the conclusion is INCONCLUSIVE and not suitable for audio forensic examination.

3. METHODOLOGY

Audio forensics uses two analytical methods to analyze the authenticity of sound recording evidence, namely Pitch analysis.

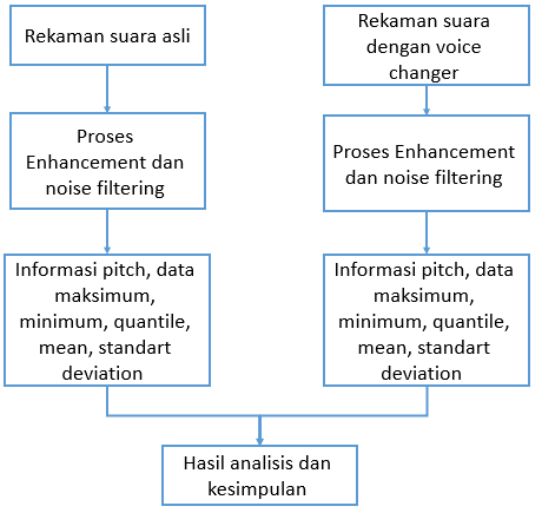


Figure 1. Research Methodology

4. DISCUSSION

In this section, we will discuss the experimental results of sound recording sample analysis. The voice sample used is the original voice recording and which has been changed using an application on an Android smartphone, namely Voice Changer With Effects. The voice recording consists of 20 words, namely "Congratulations on your mobile number being selected to win the Handa New Mobilia car prize by entering the winner's pin code RKJ789 on the Toko Prize website". The original sound is then given a telephone effect.

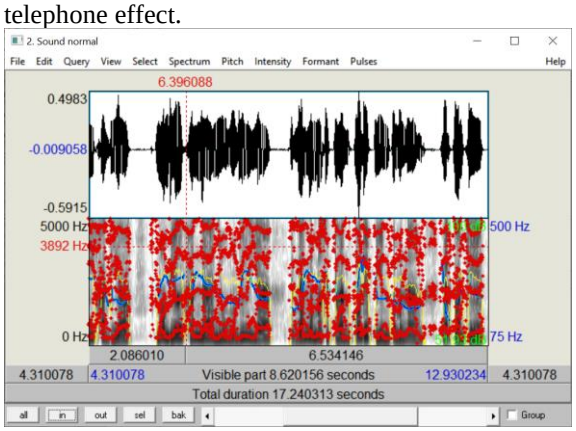


Figure 2. Original sound graphics

The red graph shows formant, the blue graph shows Pitch, and the yellow graph shows intensity, while the top window shows the frequency of uttering sentences in mono(1 channels).

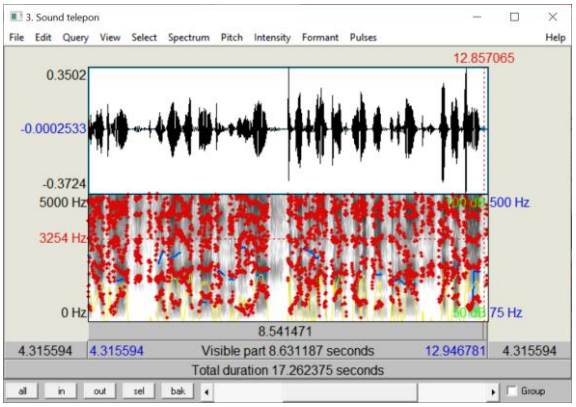


Figure 3. Voice chart with voice changer

4.A. 4.1. Statistical Pitch Analysis

This analysis is based on statistical calculations of the Pitch values of each unknown (Voices of Evidence) and known (Votes of Subject1) votes. The Pitch characteristics of each sound are compared to the minimum Pitch, maximum Pitch and mean Pitch.

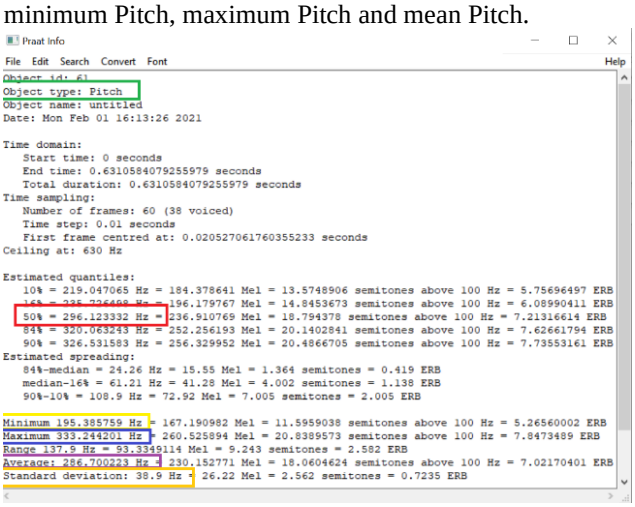


Figure 4. Retrieval of Pitch statistical values

The green box indicates the information obtained in the form of Pitch. The red box indicates the Pitch quantile value. The yellow box indicates the minimum Pitch value. The blue box indicates the maximum Pitch value. The purple box indicates the mean pitch. The orange box indicates Pitch standard deviation.

Table 1. Pitch Statistical Analysis on the pronunciation of the word congratulations

Statistical Analysis	Original Voice (Hz)	Voice Voice changer (Hz)
*pitches*minimum	195.385759	288.501169
*pitches*maximum	333.244201	619.822577
*pitches*quantile	296.123332	317.116622
*pitches*mean	286.700223	361.967927
*pitches*standard deviation	38.9	117.2

Table 2. Pitch Statistical Analysis on the pronunciation of number words

Statistical Analysis	Original Voice (Hz)	Voice Voice changer (Hz)
*pitches*minimum	121.678473	234.42138
*pitches*maximum	373.777579	342.235517
*pitches*quantile	327.060339	323.168326
*pitches*mean	297.949866	306.858128
*pitches*standard deviation	62.76	35.4

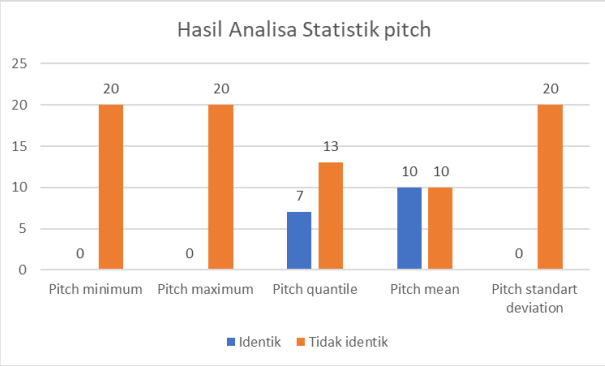
Table 3. Pitch Statistical Analysis on the pronunciation of the word handphone

Statistical Analysis	Original Voice (Hz)	Voice Voice changer (Hz)
*pitches*minimum	262.05285	293.978522
*pitches*maximum	353.850167	323.410533
*pitches*quantile	314.980437	317.023389
*pitches*mean	311.924104	312.192743
*pitches*standard deviation	12.97	9,885

To draw conclusions, the easiest and strongest argument is the mean value, followed by other statistical values.

Table 4. Results of Pitch Statistical Analysis

	Input (word)	Identical	Not identical
*pitches*minimum	20	0	20
*pitches*maximum	20	0	20
*pitches*quantile	20	7	13
*pitches*mean	20	10	10
*pitches*standard deviation	20	0	20



The pitch value is very dependent on the intonation level of the spoken voice. For example, in recording evidence, the subject speaks in the original voice, but when a comparison voice sample is taken, the subject speaks with a voice changer. From this analysis, it can be concluded that based on the statistical value, the original voice pitch is NOT IDENTICAL with the voice changer's voice

4.B. 4.2. Bandwidth Format Analysis, Graphical Distribution and Spectogram

Apart from Pitch analysis, formant bandwidth, graphical distribution and spectrogram analysis methods can also be used. Bandwidth format analysis is based on One-way Anova(Analysis of Variances) analysis which statistically calculates the Formant 1, Formant 2, Formant 3 and Formant 4 values of the original voice and the voice changer's voice. A more detailed study of the statistical analysis of formants and bandwidth is to use the Likelihood Ratio (LR) which is a continuation of the ANOVA analysis described above. This LR analysis can be used to strengthen the results of the Anova analysis obtained previously.

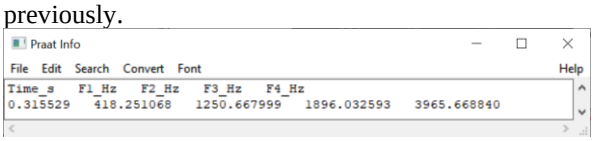


Figure 5. Original sound formant

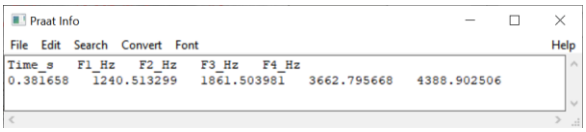


Figure 6. Voice changer sound formant

Spectrogram analysis shows a general pattern that is typical of spoken words and a special pattern that is typical of each of the syllable formants analyzed.

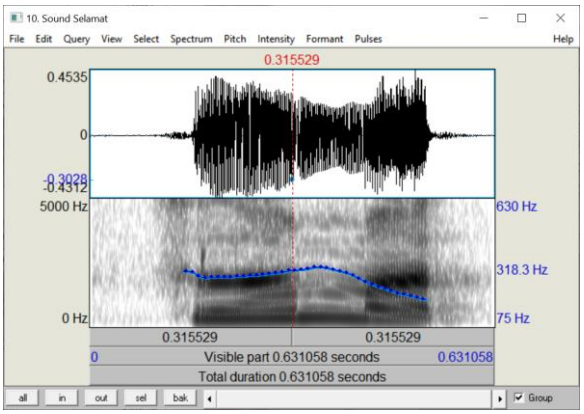


Figure 7. Screenshot of the spectrogram for the pronunciation of the word 'congratulations' from Suara Asli

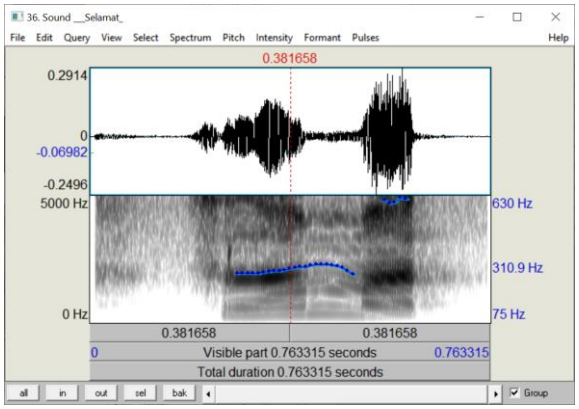


Figure 8. Screenshot of the spectrogram for the pronunciation of the word 'congratulations' from Voice changer Voice

These typical patterns for the pronunciation of certain words from unknown sounds (voices of evidence) and known (voices of comparison) do not show a significant difference, so it can be concluded that the two sounds for pronunciation of these words are IDENTICAL (have similarities spectrograms). In general, in voice recognition, the level of Pitch statistical analysis is positioned under formant and spectrogram analysis.

5. CONCLUSION

The pitch value is very dependent on the intonation level of the spoken voice. For example, in recording evidence, the subject speaks in the original voice, but when a comparison voice sample is taken, the subject speaks with a voice changer. Of the 20 spoken words, there is a difference in the value of the Pitch statistical analysis from the original voice recording to the voice given by the voice changer. So that just analyzing the Pitch value is not strong enough, an analyst is required to use the bandwidth format with the Anova and Likelihood Ratio methods, as well as graphical distribution and spectrum analysis.

BIBLIOGRAPHY

AL AZHAR, M. N. (2012). *Digial Forensic : Practical Guidelines for Computer Investigator*. Salemba Infotek.

DEVA, B. S., & MARDIANTO, I. (2019). Teknik Audio Forensik Menggunakan Metode Analisis Formant Bandwidth, *Pitch* dan Analisis Likelihood Ratio. *Ultimatics*, *10*(2), 67–72. https://doi.org/10.31937/ti.v10i2.936

HUIZEN, R. R., JAYANTI, N. K. D. A., & HOSTIADI, D. P. (2017). *Model Evaluasi Rekaman Percakapan Di Audio Forensik*.

MEHMET MEHDI KARAKOC. (2017). *Visual and Auditory Analysis Methods for Speaker Recognition in Digital Forensic*.

NIGROHO, P. D., & AL-AZHAR, MU. N. (2017). IT : DIGITAL FORENSIC. *IPSIKOM*, *5*(JUNI 2017), 1–5. http://ojs.ipem.ecampus.id/ojs_ipem/index.php/stmik-ipem/article/view/31

RUSYDI, U., SUNARDI, & GUSTAFI, M. F. (2019). Analisis Statistik Manipulasi *Pitch* Suara. *Jurnal Mobile and Forensics*, *1*(1), 1–12. https://doi.org/http://dx.doi.org/10.12928/mf.v1i1.702

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). http://ijns.org/journal/index.php/ijns/article/view/39/38

UMAR, R., RIADI, I., & HANIF, A. (2019). Analisis Bentuk Pola Suara Menggunakan Ekstraksi Ciri Mel-Frequencey Cepstral Coefficients (MFCC). *CogITo Smart Journal*, *4*(2), 294. https://doi.org/10.31154/cogito.v4i2.130.294-304