JOIN (Jurnal Online Informatika)
Volume 6 No. 2 | December 2021: 145-154
DOI: 10.15575/join.v6i2.702
p-ISSN: 2528-1682
e-ISSN: 2527-9165

The Comparison of Audio Analysis Using Audio Forensic Technique and Mel Frequency Cepstral Coefficient Method (MFCC) as the Requirement of Digital Evidence

Helmy Dzulfikar[1], Sisdarmanto Adinandra[2], Erika Ramadhani[3]
[1,2,3] Informatics Department of Post Graduate Program, Universitas Islam Indonesia, Yogyakarta

Article Info

Article history:

Received February 16, 2021
Revised April 07, 2021
Accepted May 18, 2021
Published December 26, 2021

Keywords:

Audio Forensic
Pitch Formant Spectogram
MFCC DTW KNN
Voice Identification

ABSTRACT

Audio forensics is the application of science and scientific methods in handling digital evidence in the form of audio. In this regard, the audio supports the disclosure of various criminal cases and reveals the necessary information needed in the trial process. So far, research related to audio forensics is more on human voices that are recorded directly, either by using a voice recorder or voice recordings on smartphones, which are available on Google Play services or iOS Store. This study compares the analysis of live voices (human voices) with artificial voices on Google Voice and other artificial voices. This study implements the audio forensic analysis, which involves pitch, formant, and spectrogram as the parameters. Besides, it also analyses the data by using feature extraction using the Mel Frequency Cepstral Coefficient (MFCC) method, the Dynamic Time Warping (DTW) method, and applying the K-Nearest Neighbor (KNN) algorithm. The previously made live voice recording and artificial voice are then cut into words. Then, it tests the chunk from the voice recording. The testing of audio forensic techniques with the Praat application obtained similar words between live and artificial voices and provided 40,74% accuracy of information. While the testing by using the MFCC, DTW, KNN methods with the built systems by using Matlab, obtained similar word information between live voice and artificial voice with an accuracy of 33.33%.

Corresponding Author:

Helmy Dzulfikar
Informatics Department of Post Graduate Program,
Universitas Islam Indonesia,
Jl. Kaliurang Km. 14,5 Kec. Ngemplak, Kabupaten Sleman, Daerah Istimewa Yogyakarta, Indonesia
Email: 18917112@studets.uii.ac.id

1. INTRODUCTION

The rapid development of the digital era results in an increasing number of smartphone users since it offers various facilities. In this regard, many multimedia applications on the smartphone available on the Google Play service or the iOS Store with their respective advantages can process images, sounds, and videos [1]. The large number of Android-based smartphone users enable5 both positive and negative impacts [2]. However, it is possible to misuse the sound recordings or other personal recordings on smartphones by irresponsible users for committing crimes. Therefore, many cases involve sound recording as the crucial evidence for the investigation and disclosure of those cases [3]. A trial involving digital proof requires an expert witness to assist a judge in making a verdict in a case. An expert staff supports the trial process by presenting the results of his analysis of evidence that is authentic, comprehensive, and following his scientific field through the appropriate stages and procedures [4]. The current technological advances allow the human voice to command on computer devices [5]. In this case, sound recording is metadata to get clues, such as the individual's identity, the incident location, and the time [6]. The use of sound for such evidence needs to record first, and this process goes through the scientific method to make it acceptable as evidence in a trial [7]. In this regard, the evidence used to investigate the case is analysed and read by the staging process of audio forensic [8].

Audio Forensics is a process to generate information such as features, crime scenes, and conversation transcripts [9]. According to Muhammad Nuh Al Azhar, as written in his book entitled Audio Forensics: Theory and Analysis, the parameters that can be measured and analysed include pitch, formant, and

spectrogram [10]. In this case, each person's voice varies due to many factors. It includes the differences in the vocal cords and larynx shape and size, body size, and how the articulation of a voice by a person [11]. However, so far, previous research related to audio or voice has mainly analysed the human voice. These studies include research that discusses the comparison of voice recording similarity signal [12], voice changer [13], voice verification [14] [15], voice identification [16] [17] [18], a comparison of methods, and improved sound accuracy [19] [20] to determine the characteristics of a sound. Meanwhile, this research focuses more on making comparisons between live voice (human voice) and artificial voice (google voice) to determine the similarities between the two using audio forensic techniques and several other methods. Additionally, other studies have also discussed the use of voice as a key to voice-based speech recognition [21] [22] [23] [24] [25] [26] or as a voice-based control [27] [28].

Meanwhile, the rapid development of technology leads to various recording devices for similar sound results to the actual human voice. From many studies related to sound, there is still little discussion on live voice and artificial voice. Therefore, research about live and artificial voices needs to conduct widely, and hence it can contribute positively to handling digital crimes in the future.

2. METHOD

2.1 Research Tools and Materials

This study uses some tools and materials for testing and implementing the research that supports the acquisition of the required information. The devices used in this study consisted of hardware and software. The hardware used in this study is a computer with the specifications presented in Table 1.

Table 1. Hardware

No.	Name	Specification
1.	Processor	Intel® Core™ i7-5500U CPU @ 2.40GHz
2.	Memory	8192 MB RAM
3.	Hard Disk Drive	1000 GB
4.	VGA 1 dan VGA 2	Intel HD Graphics 5500 and AMD Radeon R5 M230
5.	Smartphone	Android

.

Meanwhile, Table 2 presents the software used in this study.

Table 2. Software

No.	Name	Specification
1.	Operating System	Windows 10 Pro 64-bit
2.	Computation Software	MATLAB R2015b
3.	Spreadsheet	Gnumeric
4.	Audio Analysis	PRAAT
5.	Recording Software	Audacity

The voice material used in this study is a live voice (voice recorded by humans) and an artificial voice taken from Google Voice via Google Translate with predefined words. Live voice recording uses a recording device on a Smartphone, in which the speaker pronounces the predetermined words. In this process, the voice recording is the voice of a woman. Whereas artificial voice recording uses the set words written on Google Translate, then are played for sound playback, and are recorded using an internal recording installed previously on the laptop.

This chapter explains the research process so that the details of the sequence and steps were made systematically and could be used to model and adjust the problem. Besides, they can also analyse the research results and the difficulties faced. The steps in this research are the development and integration of the methods described by Muhammad Nur Al-Azhar following the Standard Operating Procedure for forensic analysis from the Digital Forensic Analyst Team (DFAT) [29] and using the National Institute of Standards and Technology (NIST) [30]. Figure 1 describes the procedure.

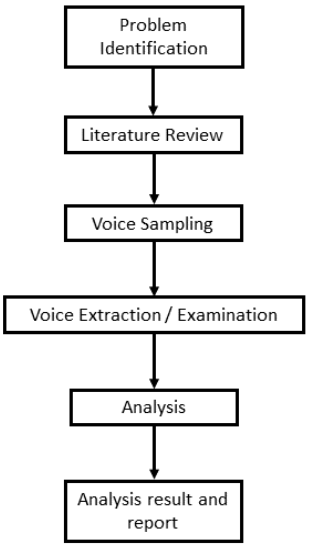


Figure 1. Research Procedure

2.2. Problem Identification

As the initial stage in this study, problem identification obtains and finds research topics to investigate further. This stage observes various phenomena, events, and information obtained from the research object in ways related to the research. Currently, problems related to voice or audio in digital crime have often occurred. However, digital crimes that use artificial voices rarely emerge. Thus, this study tries to conduct a comparative analysis of live and artificial sound. It expects to facilitate the projection of an investigation of the digital crimes that involve artificial voices as evidence.

2.3. Literature Review

The literature review collects reference materials from books, articles, papers, journals, and several sites on the internet related to the topics of this research. It includes the theory of Audio Forensic Techniques, Extraction Method with MFCC, Distance Measurement with DTW, Classification using KNN, and Analysis using Pitch, Formant, Spectrogram, and other methods to complete and present the ultimate goal of this research. The literature review is necessary to find out various discussions to make the researcher understand the extent of the investigation in this study.

2.4. Voice Sampling

This stage involves three female voices and three artificial voices (Google Voice, Responsive Voice, and Oddcast Voice) available for free. Before analysing the data, the researcher also conducts an enhancement process and noise filters to improve the quality of each voice and clean the voice from noise to get a clear voice. These processes intend to make the artificial voice taken from Google Voice and the live voice recorded by using a Smartphone have good quality and are clear from noise, so results of the analysis will not be affected.

2.5. Testing and Analysis Methods with Audio Forensic Techniques using the PRAAT Application

Figure 2 displays the manual statistical test and analysis methods using pitch, formant, and spectrogram for sound files in the form of a flowchart.

p-ISSN: 2528-1682
e-ISSN: 2527-9165

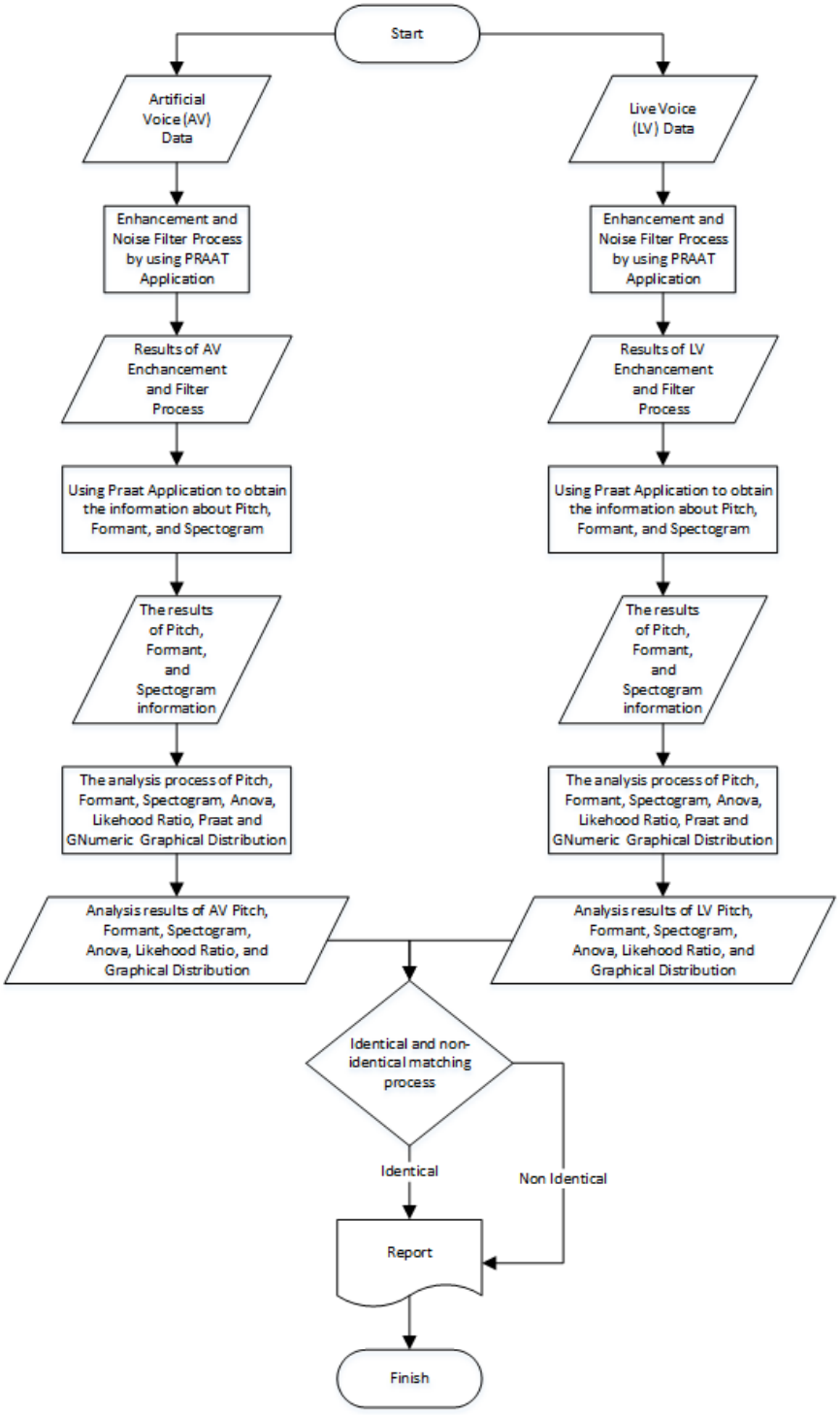


Figure 2. Testing and Analysis Flowchart using PRAAT

Muhammad Nuh Al Azhar (2011), in his book entitled *Audio Forensics: Theory and Analysis*, states that researcher can take several steps for identifying and obtaining audio information that needs some analysis techniques [10] as follow:

2.5.1 Pitch Statistical Analysis

This analysis derives from the statistical calculations of the pitch value of each unknown and known voice. The pitch characteristic of each voice compares to the minimum pitch, maximum pitch, median pitch

(quantile), mean pitch, and standard deviation pitch. If the pitch characteristics of each voice show a significant difference, then it concludes that the tone of the unknown and known voices are different.

2.5.2 Formant and Bandwith Statistical Analysis

a. Anova Analysis

ANOVA analysis shows the level of difference between 2 (two) groups of data on each formant of both unknown and known voices. It indicates the F-ratio, F-critical ratios, and P-probability. If the value of the F ratio is less than F-critical and P-value probability is more than 0.5 then the two data groups of the formant values, analysed from unknown and known voices, have no significant differences (accepted) at the 0.05 level. This conclusion has a 95% confidence rate. In concluding the Anova Formant Analysis, it requires at least formants 1, 2, and 3 to be analysed. If there are two acceptable formants within Formants 1, 2, and 3, then it is adequate to draw IDENTIC conclusions based on Anova. Nevertheless, this conclusion is also usually supported by Formant 4 or 5.

Meanwhile, some casuistic cases use bandwidth, where the subject tries to give a known voice that is significantly different aurally from the original sound, which in this case, usually uses the Pitch Shift application. Therefore in an ordinary instance, bandwidth is rarely used for voice recognition purposes.

b. Likehood Ratio Analysis (LR)

While Anova Analysis has ben explained earlier, the following is the LR formula :

$$LR = \frac{p(E \vee H_p)}{p(E \vee H_d)}$$

where:

$p(E \vee H_p)$ is the prosecution hypothesis derived from the known and unknown samples coming from the same person.

$p(E \vee H_d)$ is the defense hypothesis derived from known and unknown samples coming from different people.

$p(E|H_p)$ comes from the Anova p-value, while $p(E \vee H_d) = 1 - p(E|H_p)$

If LR > 1, then it supports $p(E|H_p)$. On the contrary, if LR < 1, then $p(E \vee H_d)$ which is supported. Thus, it is a must that the value of $p(E|H_p)$ > 0.5 to conlude that the unknown voice evidence and the known comparative voice come from the same person (IDENTIC). The following is the prosecution hypothesis as presented in Table 3.

Table 3. Verbal Statement of the Prosecution Hypothesis $p(E|H_p)$

LR	LR (log)	Verbal Statement	Explanation
> 10,000	> 4	Very strong evidence to support	Supports the Prosecution Hypothesis $p(E\|H_p)$
1,000 – 10,000	3 – 4	Strong evidence to support	
100 – 1,000	2 – 3	Moderately strong evidence to support	
10 – 100	1 – 2	Moderate evidence to support	
1 – 10	0 – 1	Limited evidence to support	

Meanwhile, the following is the defense hypothesis as presented in Table 4.

Table 4. Verbal Statement of the Defense Hypothesis $p(E \vee H_d)$

LR	LR (log)	Verbal Statement	Explanation
1 – 0.1	0 – -1	Limited evidence against	Supports the defense Hypothesis $p(E \vee H_d)$
0.1 – 0.01	-1 – -2	Moderate evidence against	
0.01 – 0.001	-2 – -3	Moderately strong evidence against	
0.001 – 0.0001	-3 – -4	Strong evidence against	
< 0.0001	> -4	Very strong evidence against	

From Table 3. and Table 4., it is discovered that to support the prosecution hypothesis (known and unknown voices coming from the same voice), it must be LR> 1, where the more the LR value, the better and stronger the verbal statement is. This LR analysis can strengthen the Anova analysis results obtained previously because this LR explains how much the LR level supports the prosecution and the defense hypothesis.

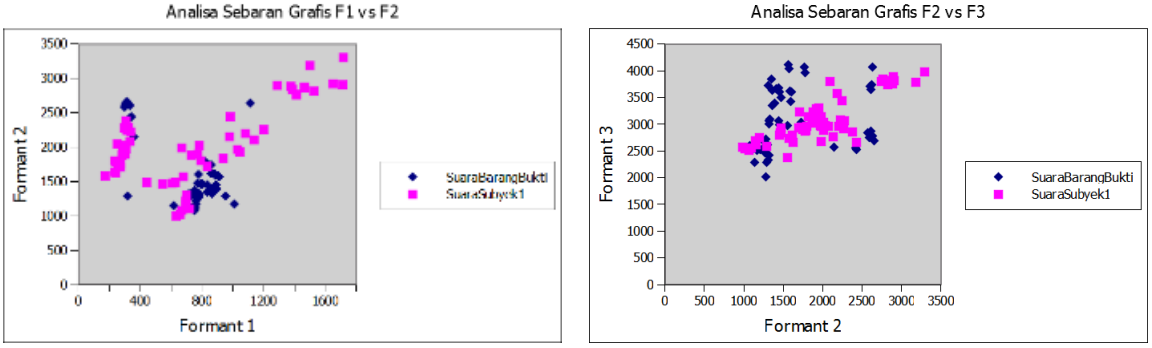
2.5.3 Graphical Distribution Analysis

The Graphical Distribution Analysis process obtained the data from the formant values extracted using PRAAT and stored in the Gnumeric application. Graphical Distribution Analysis describes the distribution

p-ISSN: 2528-1682
e-ISSN: 2527-9165

level of each formant value graphically to see the different distribution levels of formant values from unknown and known voices. In general, this analysis compares Formant 1 (F1) and Formant 2 (F2) and Formant 2 and Formant 3 (F3).

The following is an example of a pronunciation comparison of the word "*saya*" between F1 vs F2 and F2 vs F3 based on the formant values shown in Figure 3.



(a) (b)

Figure 3. (a) Graphical Distribution Analysis of F1 vs F2, (b) Graphical Distribution Analysis of F2 vs F3

The two graphs in Figure 3 illustrate several values from *SuaraSubyek1* (Subject1'sound) that come out of the group. If these values are eliminated, the graphical distribution values of F1, F2, and F3 between *SuaraBarangBukti* (EvidenceSound) and *SuaraSubyek1* (Subject1'sound) are still in the same group range (ANOVA similarity probability). In conclusion, F1, F2, and F3 between *SuaraBarangBukti* (Evidence Sound) and *SuaraSubyek1* (Subject1'sound) are IDENTIC.

2.5.4 Spectrogram Analysis

This analysis shows a regular pattern of the spoken words and a specific pattern of each analysed syllable formant. These typical patterns are included in the examination of each formant's energy level. In this case, if the typical patterns for the pronunciation of certain words from the unknown and known voice do not show a significant difference, then the two voices are IDENTIC (have the same spectrogram).

2.6. Testing and Analysis Method with Combination Method by using MATLAB Application

2.6.1 Mel Frequency Cepstrum Coefficients (MFCC)

MFCC is a widely used method in speech technology, both in speaker and speech recognition. This method performs feature extraction, which is a process that converts a voice signal into several parameters. Meri Susanti et al, in 2018 write in their journal that there are several advantages of this method, such as [31]:

a. Capturing voice characteristics that are very important for speech recognition, or in other words, it can capture crucial information contained in voice signals.
b. Producing minimum data without losing crucial information it contains.
c. Duplicating humans' hearing organs while they perceive voice signals.

The method of testing and analysis using MFCC for sound files can be displayed in a flowchart as presented in Figure 4.

2.6.2 Dynamic Time Warping (DTW) and K-Nearest Neighbour (KNN)

DTW and KNN are classification methods that function to classify the suitability of the test data with the collected data. DTW projects the matrix from the training data towards the test data with the euclidean distance equation, then adds the values diagonally by selecting the minimum value for each index shift of the matrix [18]. Meanwhile, KNN will sort or find the closest class and adjust it to the actual one.

Figure 5 presents the testing and analysis methods using DTW – KNN for the sound files in a flowchart.

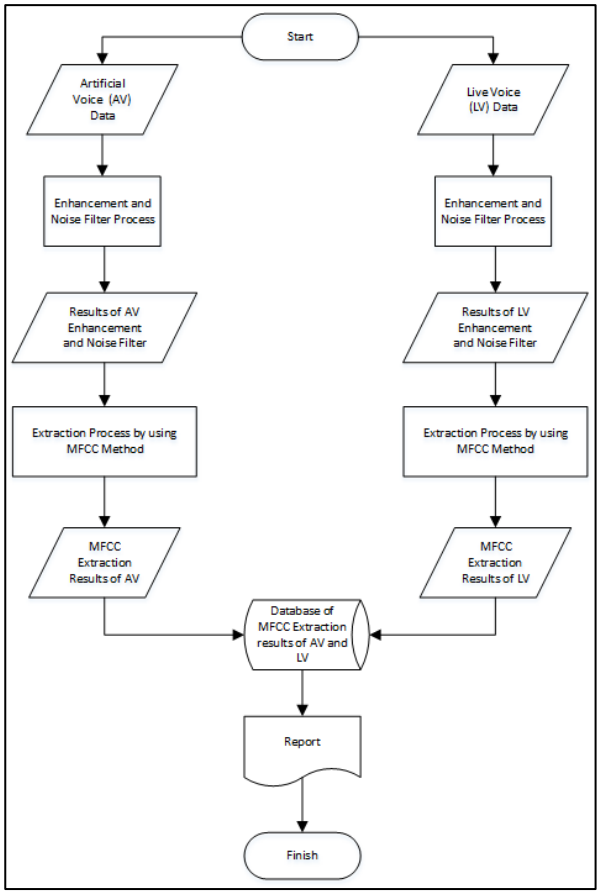


Figure 4. MFCC Testing Flowchart

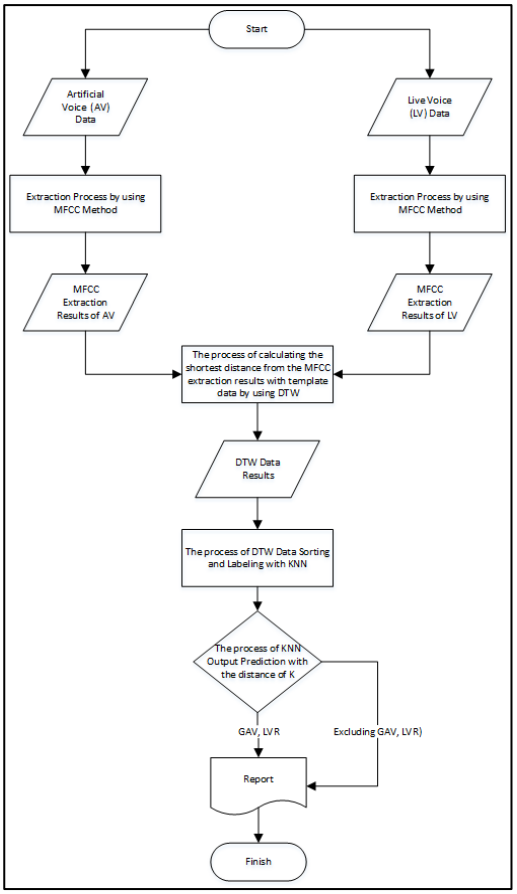


Figure 5. DTW-KNN Flowchart

3. RESULTS AND DISCUSSION

The process of analysing the implementation of existing methods requires an experimental case scenario involving the recording of the original voice (unknown) and the recording of the comparator voice (known). This case scenario experiment analysed Sound recording evidence. The decoding stage obtained a complete transcript of the recorded conversation voice as a suspect. Furthermore, the recording transcript is analysed to determine whether the suspect voice is identical to the comparator voice. It requires getting 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 evidence is Identic with a comparator voice. It refers to "Speaker Voice Identification: A Forensic Survey" written by Koenig, B.E. from the Federal Bureau of Investigation (FBI) [10] [32]. The following is the voice recording used in this study, in which its sentences have been broken down into words namely, *"silahkan", "kamu", "transfer", "dulu", "ke", "rekening", "aku", "sebesar", "dua", "juta", "rupiah", "nanti", "nomor", "kamar", "dan", "nama", "hotelnya", "aku", "kirim", "lewat", "whatsapp", "setelah", "transfer", "aku", "cinta", "kamu", "mas"*. The sentence is broken down into 27 words dan 22 words that have different meanings.

3.1. Test and Analysis Results with Audio Forensic Techniques by using the PRAAT Application

Table 5, Table 6, and Table 7 present the summary of the analysis towards the tested data using PRAAT as follow:

Table 5. Testing Results of Google Artificial Voice (GAV) and Live Voice Recording (LVR)

Methods	Training Data (SUM)	Test Data (SUM)	Results Detected (SUM)		Accuracy
			GAV	LVR	
Pitch Analysis	27 words (gav)	27 words (lvr)	6 words	21 words	77,78 %
Anova Analysis	27 words (gav)	27 words (lvr)	3 words	24 words	88, 89 %
Likehood Ratio (LR) Analysis	27 words (gav)	27 words (lvr)	21 words	6 words	22,22 %
Graphical Distribution Analysis	27 words (gav)	27 words (lvr)	3 words	24 words	88,89 %
Spectogram Analysis	27 words (gav)	27 words (lvr)	6 words	21 words	77,78 %
Accuracy Total					71,11 %

Table 6. Testing Results of Google Artifical Voice (GAV) and Artificial Voice of Responsive Voice (AVR)

p-ISSN: 2528-1682
e-ISSN: 2527-9165

Methods	Training Data (SUM)	Test Data (SUM)	Results Detected (SUM)		Accuracy
			GAV	AVR	
Pitch Analysis	27 words (gav)	27 words (avr)	22	5 words	18,52 %
Anova Analysis	27 words (gav)	27 words (avr)	20	7 words	25,93 %
Likehood Ratio (LR) Analysis	27 words (gav)	27 words (avr)	26	1 words	3,70 %
Graphical Distribution Analysis	27 words (gav)	27 words (avr)	26	1 words	3,70 %
Spectogram Analysis	27 words (gav)	27 words (avr)	27	0 words	0,0 %
Accuracy Total					10,37 %

Table 7. Testing Results of Artificial Voice and Live Voice by using Audio Forensic Technique

Methods	Training Data (sum)	Tested Data (sum)	Accuracy
Result of Accuracy in Table 3.	27 words (gav)	27 words (lvr)	71,11 %
Result of Accuracy in Table 4.	27 words (gav)	27 words (avr)	10,37 %
Accuracy Total			40,74 %

The testings results show that the level of similarity between Google artificial voice recordings (gav) and live voice recordings (lvr) as presented in Table 3., are different, seen from all forensic audio analyses that have been carried out. Table 3. shows that the level of accuracy is 71.11%. Meanwhile, the similarity level test was carried out between the artificial voice recording by Google Voice (gav) and the artificial voice recording by responsive voice (avr), which its results are presented in Table 4. In Table 4. It can be seen that an accuracy rate is 10.37%, while the accumulation of test and analysis results by using audio forensic techniques obtained an accuracy of 40,74%.

In this case, the Pitch value is very influential on a voice, as written in a study [13], since the performance of a voice conversion system is affected by alpha scale (α) factors, and beta (β) scale factors. During the Time Domain Pitch Synchronous Overlap Add (TD-PSOLA) process, the scale factors of alpha (α) and beta (β) were used for time stretching and pitch shifting parameters. In this regard, the more alpha (α) and beta (β) pause approach 1 (one), the voice conversion result will sound more similar to the original / input voice, and vice versa.

3.2. The Test and Analysis Results with Audio Forensic Techniques by using the MATLAB Application

This research used 27 training voices in the database to process the test of the input test voice. Each class was tested 27 times using 54 sounds precluded from the training voice. Those voices split into two, namely one different artificial voice recording and one different live recording, in which each voice recording consists of 27 words.

In addition, this process introduced a new input voice. The following is the process for the test voice:

a. Sounds from the artificial voice recordings or live voice recordings used as test sounds obtained from the recording process or the voices selected from the live directory. The sounds are in the form of (.wav) with a sampling frequency of 16000 Hz.
b. The next step is importing the voice into the system.
c. Then, MFCC is used in the feature extraction stage of the tested data.
d. The MFCC process on the test voice obtains several parameters to analyse in the form of a feature vector paired with the feature extraction results of the training voice.
e. The characteristic vector of the input voice paired with the database that has been stored previously in the training voice database. The matching process uses DTW, in which the test voice vector is matched with the training voice vector to obtain the minimum value, indicating the kind of sound.

The KNN process is conducted based on the result of the DTW process. The obtained results after the test are in the form of class recognition result (word prediction) and the value of similarity measurement values in the form of proportions for the KNN method and for the DTW method, at which the smaller the number, the more similar it will be. Then each recognition result will be recapitulated and the accuracy value measured based on the respective classification method.

Meanwhile, the summary results of testing the artificial voice and the live voice is shown in Table 6.

Table 8. Test Results of Google Artificial Voice (GAV) and Live Voice Recording (LVR)

Methods	Training Data (SUM)	Test Data (SUM)	Detected Results (SUM)		Accuracy
			GAV	LVR	
MFCC, DTW and KNN	27 words (gav)	27 words (lvr)	27 word	0 words	00,00%
	27 words (gav)	27 words (rav)	9 words	18 words	66,67%
Accuracy Total					33,33%

In conclusion, based on the test results in Table 6, the tested words include 54 words of artificial voice recordings and new live voice recordings, in which those words are different from the words in the training data.

The test data, words from other artificial sound recordings, indicate that some are IDENTICAL to the artificial voice recordings. Unlike the case with live voice recordings, the results were NOT IDENTICAL to the artificial sound in the training data. Therefore, the similarity level trials conducted between Google artificial voice recordings (GAV) with live voice recordings (LVR) and Google artificial voice recordings (GAV) with Responsive artificial voice recordings (AVR) as presented in Table 5 show an accuracy of 33.33%.

4. CONCLUSION

The identification results of voice comparisons using audio forensic techniques are obtained from predetermined stages and procedures. In this regard, the audio forensic analysis technique is more effective since it shows the level of similarity between artificial voice recordings and live voice recordings with a 40.74% accuracy value. However, the audio forensic technique is less efficient. In this case, the process of audio forensic analysis takes a long time. It is affected by the length of time and the quality of the two voice recordings. Contrarily, comparing the similarity level between Live Voice and Artificial Voice identification using a built system employing the MFCC extraction method, matching with DTW, and classification with KNN is quite efficient in analysing voice recordings. The analysis results are less effective because it is affected by the accuracy of determining parameters used as a measurement in the built system, of which the obtained accuracy is 33.33%. The analysis of artificial and live voices are more effective using audio forensic technique than the built system. Nonetheless, in terms of completion time, it is more efficient to use a built system for an investigation process that requires the results of an evidence item to be delivered faster at the initial trial. To sum up, the results show that the two voices have very different characteristics from their pitch values.

5. REFERENCES

[1] A. Subki, "Suara Voice Changer dengan Rekaman Suara," Univeritas Islam Indonesia, Yogyakarta, 2017.

[2] I. Riadi, A. Yudhana, and M. C. F. Putra, "Forensic Tool Comparison on Instagram Digital Evidence Based on Android with The NIST Method," *Sci. J. Informatics*, vol. 5, no. 2, pp. 235–247, 2018, doi: 10.15294/sji.v5i2.16545.

[3] G. Wicaksono and Y. Prayudi, "Teknik Forensika Audio Untuk Analisa Suara Pada Barang Bukti Digital," *Semnas Unjani*. Semnas Unjani, Bandung, pp. 1–6, 2013.

[4] V. A. H. Firdaus, "Forensik audio pada rekaman suara," *Sch. Electr. Eng. Informatics Inst. Technol. Bandung Bandung, Indones.*, p. 1, 2016.

[5] R. Umar, I. Riadi, and A. Hanif, "Analisis Bentuk Pola Suara Menggunakan Ekstraksi Ciri Mel-Frequencey Cepstral Coefficients (MFCC)," *CogITo Smart J.*, vol. 4, no. 2, p. 294, 2019, doi: 10.31154/cogito.v4i2.130.294-304.

[6] H. Malik, "Acoustic environment identification and its applications to audio forensics," *IEEE Trans. Inf. Forensics Secur.*, vol. 8, no. 11, pp. 1827–1837, 2013, doi: 10.1109/TIFS.2013.2280888.

[7] H. Zhao and H. Malik, "Audio recording location identification using acoustic environment signature," *IEEE Trans. Inf. Forensics Secur.*, vol. 8, no. 11, pp. 1746–1759, 2013, doi: 10.1109/TIFS.2013.2278843.

[8] A. P. Saputra, H. Mubarok, and N. Widiyasono, "Analisis Digital Forensik pada File Steganography (Studi kasus : Peredaran Narkoba)," *Tek. Inform. dan Sist. Inf.*, vol. 3, no. 1, pp. 179–190, 2009.

[9] R. R. Huizen, N. K. D. A. Jayanti, and D. P. Hostiadi, "Analisis Pengaruh Sampling Rate Dalam Melakukan Identifikasi Pembicara Pada Rekaman Audio," *Konferensi Nasional Sistem & Informatika*. pp. 9–10, 2015.

[10] M. Al-Azhar Nuh, "Audio Forensics : Theory and Analysis," pp. 1–38, 2011.

[11] D. K. Putra, I. Iwut, and R. D. Atmaja, "Simulasi Dan Analisis Speaker Recognition Menggunakan Metode Mel Frequency Cepstrum Coefficient (MFCC) Dan Gaussian Mixture Model (GMM)," *eProceedings Eng.*, vol. 4, no. 2, pp. 1766–1772, 2017.

[12] A. Aligarh and B. C. Hidayanto, "Implementasi Metode Forensik dengan Menggunakan Pitch, Formant, dan Spectrogram untuk Analisis Kemiripan Suara Melalui Perekam Suara Telepon Genggam Pada Lingkungan yang Bervariasi," *J. Tek. ITS*, vol. 5, no. 2, 2016, doi: 10.12962/j23373539.v5i2.16980.

[13] A. Subki, B. Sugiantoro, and Y. Prayudi, "Analisis Rekaman Suara Voice Changer dan Rekaman Suara Asli Menggunakan Metode Audio Forensik," *Indones. J. Netw. Secur.*, vol. 7, no. 1, p. 1, 2018, [Online]. Available: http://ijns.org/journal/index.php/ijns/article/view/39/38.

[14] A. Kurniawan, "Verifikasi Suara menggunakan Jaringan Syaraf Tiruan dan Ekstraksi Ciri Mel Frequency Cepstral Coefficient," *J. Sist. Inf. Bisnis*, vol. 7, no. 1, p. 32, 2017, doi: 10.21456/vol7iss1pp32-38.
[15] Hans Kalveram and Peter Meissner, "Itakura-saito clustering and rate distortion functions for a composite source model of speech," vol. 18, pp. 195–216, 1989.
[16] B. S. Deva and I. Mardianto, "Teknik Audio Forensik Menggunakan Metode Analisis Formant Bandwidth, Pitch dan Analisis Likelihood Ratio," *Ultimatics*, vol. 10, no. 2, pp. 67–72, 2019, doi: 10.31937/ti.v10i2.936.
[17] U. Rusydi, "Analisis Statistik Manipulasi Pitch Suara," *J. Mob. Forensics*, vol. 1, no. 1, pp. 1–12, 2019, [Online]. Available: http://journal2.uad.ac.id/index.php/mf/article/view/702.
[18] V. R. C. Putri and Sunarno, "Analisis Rekaman Suara Menggunakan Teknik Audio Forensik Untuk Keperluan Barang Bukti Digital," *Unnes Phys. J.*, vol. 3, no. 1, 2014.
[19] A. Irawan, "Perbandingan Metode Itakura-Saito Distance dan Manual Statistik (Pitch, Formant, Spectogram) untuk Akurasi Identifikasi Suara pada Audio Forensik," Universitas Islam Indonesia, 2019.
[20] A. Wicaksono, S. Adinandra, and Y. Prayudi, "Penggabungan Metode Itakura Saito Distance dan Backpropagation Neural Network untuk Peningkatan Akurasi Suara pada Audio Forensik (Combining Itakura Saito Distance and Backpropagation Neural Network Methods to Improve Sound Accuracy in Audio Forensic)," *JUITA - J. Inform.*, vol. 8, no. November, pp. 225–233, 2020.
[21] M. N. Rabbani, A. Rizal, and F. Y. Suratman, "Implementasi Kunci Berbasis Suara Menggunakan Metode Mel Frequency Frequency Cepstral Coefficient (MFCC)," vol. 3, no. 3, pp. 3998–4007, 2016.
[22] R. A. Sadewa, T. A. B. Wirayuda, and S. Sa'adah, "Implementasi Speaker Recognition Untuk Otentikasi Menggunakan Modified Mfcc–Vector Quantization Algoritma LBG," *eProceedings Eng.*, vol. 2, no. 1, pp. 1453–1463, 2015.
[23] F. Elkusnandi, Adiwijaya, and U. N. Wisesty, "Implementasi Sistem Pengenalan Ucapan Bahasa Indonesia Menggunakan Kombinasi MFCC dan PCA Berbasis HMM," vol. 5, no. 2, pp. 3608–3622, 2018.
[24] I. S. Permana, Y. Indrawaty, and A. Zulkarnain, "Implementasi Metode Mfcc Dan Dtw Untuk Pengenalan Jenis Suara Pria Dan Wanita," *MIND J.*, vol. 3, no. 1, pp. 61–76, 2019, doi: 10.26760/mindjournal.v3i1.61-76.
[25] A. Setiawan, A. Hidayatno, and R. R. Isnanto, "Aplikasi Pengenalan Ucapan dengan Ekstraksi Mel-Frequency Cepstrum Coefficients (MFCC) Melalui Jaringan Syaraf Tiruan (JST) Learning Vector Quantization (LVQ) untuk Mengoperasikan Kursor Komputer," *Apl. Pengenalan Ucapan dengan Ekstraksi Mel-Frequency Cepstrum Coefficients Melalui Jar. Syaraf Tiruan Learn. Vector Quantization untuk Mengoperasikan Kursor Komput.*, vol. 13, no. 3, pp. 82–86, 2011, doi: 10.12777/transmisi.13.3.82-86.
[26] R. Umar, I. Riadi, A. Hanif, and S. Helmiyah, "Identification of speaker recognition for audio forensic using k-nearest neighbor," *Int. J. Sci. Technol. Res.*, vol. 8, no. 11, pp. 3846–3850, 2019.
[27] M. Azizah, A. Hidayatno, and Y. Christyono, "Aplikasi Pengenal Pengucap Berbasis Identifikasi Suara Dengan Ekstraksi Ciri Mel-Frequency Cepstrum Coefficients (Mfcc) dan Kuantisasi Vektor" *Transient*, vol. 6, no. 4, pp. 639–643, 2017.
[28] I. Fadli, "Kendali Pintu Air Otomatis Berbasis Speech Recognition Menggunakan Metode MFCC dan Jaringan Syaraf Tiruan," vol. XV, no. 2, pp. 56–61, 2016.
[29] M. Al-Azhar Nuh, *Digital Forensics Practical Guidlines for Computer Investigation*. Jakarta, 2012.
[30] Mustafa, I. Riadi, and R. Umar, "Rancangan Investigasi Forensik E-mail dengan Metode National Institute of Standards and Technology (NIST)," *Snst Ke-9*, vol. 9, pp. 121–124, 2018.
[31] M. Susanti, B. Susilo, and D. Andreswari, "Aplikasi Speech-To-Text Dengan Metode Mel Frequency Cepstral Coefficient (MFCC) Dan Hidden Markov Model (HMM) Dalam Pencarian Kode," *J. Rekursif*, vol. 6, no. 1, pp. 48–58, 2018.
[32] B. E. Koenig, "Spectrographic voice identification: A forensic survey," *J. Acoust. Soc. Am.*, vol. 79, no. 6, pp. 2088–2090, 1986, doi: 10.1121/1.393170.