Forensic Analysis of AI-Modulated Threatening Voices: A PRAAT-based Study

G KRISHNA[1], KAJAL BANSAL[2]
[1] Post Graduate Student, Department of Forensic Science, Garden City University, Bangalore
[2] Assistant Professor Department of Forensic Science, Garden City University, Bangalore

Abstract- ***Audio analysis is defined as the systematic investigation of recorded audio evidence to determine its origin, content, and authenticity. It plays a vital role in forensic investigations, especially in analysing audio recordings to determine the identity of speakers through comparison of known and unknown samples. Audio comparison involves examining a variety of auditory parameters such as pitch, intensity, jitter, pulse, duration, and spectral characteristics. This investigation seeks to determine the likelihood of a link between recognized and questioned voices. This study examines audio forensics involving AI-altered threatening voices, addressing concerns regarding the potential harm posed by AI's use of voice-altering software to create threatening messages. This study explores a technique called PRAAT analysis to determine if it can effectively identify AI-manipulated voices used for malicious intent. Using a careful methodology, audio samples will be collected from participants acting out threatening scenarios based on prepared prompts. Then, these samples will be altered using software called VOICEMOD CLIPS. The PRAAT software will be used to analyse the original human voice and robotic voices which is modulated using AI, focusing on factors such as pitch, formant frequencies, jitter, pulse and spectrographic representations. By looking at these aspects along with ways to identify speakers and conducting a thorough analysis, the study aims to give useful insights for law enforcement and security agencies dealing with technology-based threat detection.***

Index Terms- ***AI-modulated voices, voice forensics, PRAAT analysis, threat detection, voice analysis, speaker identification, voice manipulation.***

I. INTRODUCTION

Audio analysis involves the systematic examination and interpretation of sound recordings. It encompasses the application of scientific principles and specialized techniques to analyse the characteristics, content, and context of audio recordings, aiming to extract relevant information that can assist in legal investigations or proceedings. Forensic audio analysis encompasses a wide range of activities, including Authentication, Voice Identification, Speaker Profiling, Transcription and Interpretation, Enhancement, Speaker Attribution, Stress Analysis, and Forensic Phonetics. Audio plays a pivotal role as evidence in forensic investigations due to its unique ability to provide insights into identity, behaviour, and intent. Through the analysis of vocal characteristics such as pitch, timbre, and accent, voice recordings can be used to identify individuals and corroborate testimonies, thereby strengthening the credibility of evidence presented in court. Furthermore, recent studies, such as the research by Magdin et al. (*Voice Analysis Using PRAAT Software and Classification of User Emotional State*), have demonstrated the effectiveness of PRAAT software in analysing voice recordings for emotional state classification, further highlighting the utility of advanced tools in deciphering the intricate emotional cues embedded within voice recordings. Whether it's verifying alibis, assessing threats, or discerning motives, voice evidence enables forensic experts to reconstruct events, evaluate testimonies, and support legal proceedings with objective and tangible proof.

Artificial Intelligence (AI) has become an integral part of daily life, offering convenience and efficiency across various domains. From smart assistants like Siri and Alexa aiding in task management to chatgpt and Gemini which helps in answering queries, its role in simplifying tasks, personalizing experiences, and improving efficiency in daily life. However, recent developments have raised concerns regarding AI's potential negative impacts, such as the manipulation of voice modulation technology for malicious purposes, as highlighted in studies like the one conducted by Kuo-Liang Huang et al. (*Affective Voice Interaction and Artificial Intelligence*). These advancements in AI have led to the emergence of voice modulation technology, which, while initially designed for positive applications like enhancing speech synthesis and accessibility, unfortunately, can be manipulated for nefarious purposes. While Artificial Intelligence (AI) offers numerous benefits and

advancements, it also presents significant concerns and challenges. One particularly troubling aspect is the misuse of AI-driven voice modulation technology. Voice modulation technology, initially designed for positive applications like enhancing speech synthesis and accessibility, unfortunately, can be manipulated for malicious purposes. Voice modulation technology, which alters the way someone's voice sounds by tweaking factors like pitch and intensity, poses a significant concern due to its potential misuse. There's a risk that individuals with ill intentions might utilize this technology to morph their voices into robotic or alien-like tones, intending to intimidate or deceive others. By obscuring their true identity through voice alteration, these individuals could issue threats or manipulate people without fear of being recognized or held accountable for their actions. Moreover, they could exploit this technology to create synthetic voices that closely mimic real individuals, making their threats appear more convincing and compelling. This highlights the importance of understanding the capabilities of voice modulation technology and implementing measures to safeguard against its misuse, ensuring the protection of individuals from coercion and manipulation through altered voices.

In criminal investigations, voice analysis has emerged as a critical tool for law enforcement agencies in identifying suspects, corroborating testimonies, and establishing guilt or innocence. Real-life cases abound with examples of how voice analysis has been instrumental in unravelling complex criminal mysteries. For instance, in cases involving extortion, ransom demands, or threatening phone calls, forensic experts can analyse the voice recordings to determine the identity of the perpetrator, providing crucial leads for investigators. Similarly, in cases of kidnapping or missing persons, voice analysis can help authenticate ransom demands or decipher distress calls, aiding law enforcement agencies in their search and rescue efforts.

By comparing the voice samples obtained from the recorded communications with known reference samples, forensic experts can establish a definitive link between the perpetrator and the harassing messages, providing compelling evidence for prosecution.

PRAAT, conceived in the academic corridors of the University of Amsterdam during the late 1980s, emerged as a pioneering endeavour by Paul Boersma and David Weenink, aimed at Exploring and figuring out the detailed rhythms and sounds in the Dutch language. Stemming from a necessity to Examine the fine details of how people speak, focusing on the rhythm and the patterns in how the tone of their voice changes. PRAAT gradually metamorphosed into a multifaceted software application, replete with an arsenal of tools for the analysis, synthesis, and manipulation of speech signals. Its nomenclature, "PRAAT," an embodiment of the Dutch word for "talk" or "speak," encapsulates its fundamental purpose—to provide a platform for the investigation of human speech.

Through decades of development, PRAAT evolved exponentially, expanding it's a collection that includes a wide range of important features necessary for researchers in various fields. From spectrogram analysis, pitch and intensity measurement, to formant analysis and phonetic transcription.

Today, PRAAT represents more than just a tool for speech analysis. It stands as a symbol of the collaborative efforts of researchers worldwide, as they work together to unravel the intricacies of human communication. Its lasting impact reflects the collaborative essence of scientific exploration, capturing the relentless pursuit of knowledge and comprehension within the ever-expanding domain of human cognition and expression.

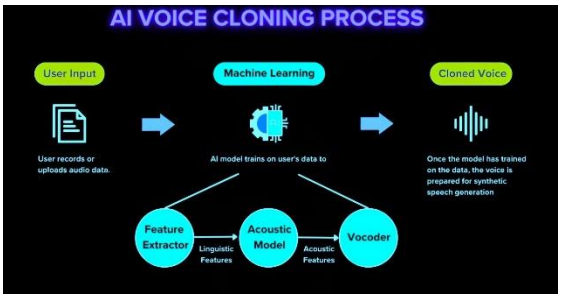

Fig. 1 Showing Voice Cloning Working Mechanism

VOICEMOD CLIPS represents a cutting-edge fusion of intuitive user interface design and sophisticated artificial intelligence algorithms, converging to revolutionize the realm of voice modulation. At its core, the application offers an expansive array of voice modulation options, spanning from subtle pitch adjustments to radical

transformations like robotic or gender-swapped voices. What sets VOICEMOD CLIPS apart is its seamless integration of AI technologies, which underpin the nuanced manipulation of voice characteristics with remarkable precision. Through the utilization of advanced machine learning algorithms, VOICEMOD CLIPS can analyse and adapt to individual vocal nuances, allowing for personalized and realistic voice alterations. Moreover, the incorporation of real-time feedback mechanisms powered by AI empowers users to preview and adjust voice modifications instantaneously, refining their creations with unparalleled accuracy. Whether it's tweaking pitch, accentuating certain tones, or completely transforming vocal identity, the intuitive interface coupled with AI assistance streamlines the entire modulation process, making it accessible to both novice users and seasoned professionals. This dynamic interaction between user input and AI-driven assistance fosters a collaborative and adaptive environment, where users can continually explore new possibilities and push the boundaries of voice modulation creativity. In the broader landscape of digital content creation, VOICEMOD CLIPS emerges as a transformative tool, empowering individuals to craft compelling audio narratives and immersive experiences with unprecedented ease and sophistication.

II. AIM OF RESEARCH

This research aims to investigate the detectability of AI-altered voices used for threatening purposes through acoustic analysis using software like "PRAAT." It seeks to identify unique acoustic features distinguishing AI-altered voices from natural human voices, potentially contributing to the prevention of harassment and nefarious activities involving AI technology. By formulating a methodological framework for detection, the study aims to enhance security in communication platforms. Through these efforts, the research aims to make society safer by protecting against potential risks from the misuse of AI-altered voices in both online and offline settings.

III. METHODOLOGY

In this study the sampling collection method are as follows: -

- 3.1 Sample Collection

Collecting voice samples from participants, disregarding gender and age.

Here 150 samples have been collected irrespective of age and gender 79 male voice samples and 70 female voice sample.

In terms of collecting the voice samples of participants script has been created which is like of threatening to another person Utilized a scripted scenario for the threatening call, aiming to evoke a natural response from participants.

In that case, utilizing a scripted scenario involves providing participants with a written script containing the threatening message, and then instructing them to read it aloud during the recording process.

This scripted scenario aims to evoke a natural response from participants as they verbalize the threatening message, simulating a realistic situation where they are delivering the message directly to someone else. The intention is to capture their authentic vocal and emotional reactions as they engage with the script, providing valuable data for analysis in this research on speaker identification and voice modulation.

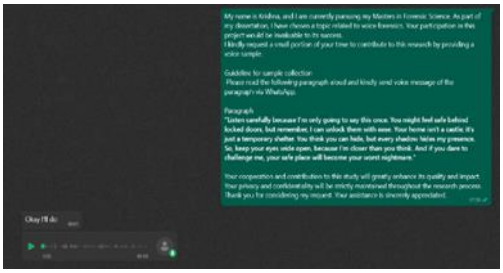

Fig. 2 Collecting voice sample with participants concern

- 3.2 Sample Modulation

After collecting the samples, a copy of the original voice sample is created for the purpose of modulating it into a robotic voice

The copied voice sample which was created earlier should be uploaded to the VOICEMOD CLIPS app from a mobile device in order to modulate the original voice into robotic voice

After uploading the voice, search for the desired robotic voice to modulate the original voice sample

Fig. 3 Showing the main interface

Fig. 4 Interface of robotic voice

After converting the copied voice sample into a robotic one, ensure to verify the audio format of the modulated voice. Because PRAAT can able to read WAV (Waveform Audio File Format), AIFF (Audio Interchange File Format) or AIFC (Audio Interchange File Format Compressed), NeXT/Sun(.au) (NeXT/Sun Audio File Format), FLAC (Free Lossless Audio Codec) and MP3 (MPEG-1 Audio Layer 3) formats

Adjust the conversion format accordingly to ensure compatibility with the PRAAT software for reading The format conversion to MP3 is being implemented here

In this research, the MP3 format is selected as it is commonly for all audio files

- 3.3 Analysis of Sample

After completing the conversions, the analysis should commence using PRAAT software.

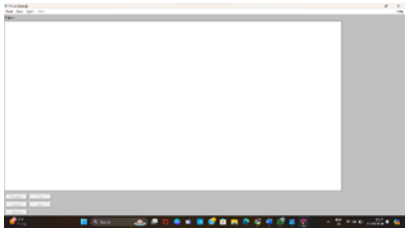

Fig. 5 This is the main interface of the PRAAT software

To begin analysis in PRAAT software, launch the program. You'll see two tabs: '*PRAAT Object'* and '*PRAAT Picture'*. Select the *'PRAAT Object'* tab.

To start the analysis, upload both the original and modulated voice samples of the same participant. You can do this by either opening the files manually or pressing '*Ctrl + O'*. After uploading the samples, various options will appear. Among these, select the 'View & Edit' option located on the right side of the PRAAT object

After uploading the samples, open each of them in separate tabs to facilitate comparative analysis

Choose the sample with specific timings for analysis and click *select*

At the top of the software interface, there are several options. Select *'Pulses'*, and within that, choose *'Voice Report'*.

Proceed to note down the values of the voice sample for the selected parameters of acoustics.

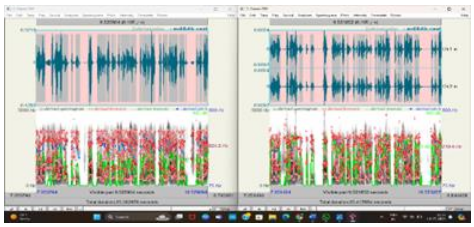

Fig. 6 Showing the analysis of sample

Fig. 7 Showing the report format of voice sample

SI No.	Parameter	Male	Female
1	Sample Size	79	71
2	Voice sample Format	MP3	MP3

Table 1: Showing the parameters of voice sample

IV. RESULT AND DISCUSSION

Parameter	Observation
Average of Mean pitch (Hz)	173.08843333333
Average of Mean pitch (Hz)	176.59684
Average of Standard deviation (Hz)	28.766873333333
Average of Minimum pitch (Hz)	107.61036666667
Average of Maximum pitch (Hz)	275.76607333333
Average of Number of pulses	888.46666666667
Average of Number of periods	859.96666666667
Average of Standard deviation of period	0.92314825333333
Average of Number of voice breaks	25.426666666667
Average of Fraction of locally unvoiced frames	44.19663%(423/955.21)
Average of Jitter (local, absolute)	144.76038
Average of Mean autocorrelation	0.92929982
Average of Mean noise-to-harmonics ratio	0.0874859
Average of Mean harmonics-to-noise ratio	14.629765453333

Table 2: Showing the mean value of original voice sample

Parameter	Observation
Average of Mean pitch (Hz)	184.1702267
Average of Mean pitch (Hz)	188.87587333333
Average of Standard deviation (Hz)	32.234486666667
Average of Minimum pitch (Hz)	129.47922
Average of Maximum pitch (Hz)	256.6441
Average of Number of pulses	560.80666666667
Average of Number of periods	507.4466667
Average of Standard deviation of period	0.968053436
Average of Number of voice breaks	16.3
Average of Fraction of locally unvoiced frames	61.6908%(589.3/955.6)
Average of Jitter (local, absolute)	289.0439
Average of Mean autocorrelation	1.42629054
Average of Mean noise-to-harmonics ratio	0.36271658
Average of Mean harmonics-to-noise ratio	5.24349258

Table 3: Showing the mean value of robotic voice sample

Interpretation

Median Pitch Difference: Robotic voices have a higher median pitch (184.17 Hz) compared to human voices (173.09 Hz), indicating that the modulation process raises the overall pitch by 11.08 Hz. This results in a less natural and more artificial sound, potentially affecting speaker identification.

Mean Pitch Difference: Robotic voices have a higher mean pitch (188.88 Hz) compared to human voices (176.60 Hz), showing an increase of 12.28 Hz. This consistent elevation impacts the perceived

identity, making the voice sound more synthetic and challenging to match with the original human voice. Pitch Variability Increase: Robotic voices exhibit greater pitch variability (32.23 Hz) compared to human voices (28.77 Hz), an increase of 3.47 Hz. This contributes to a more mechanical sound as the modulation process introduces irregular pitch changes to simulate a robotic effect.

Minimum Pitch Rise: The lowest pitch in robotic voices (129.48 Hz) is significantly higher than in human voices (107.61 Hz), a difference of 21.87 Hz. This suggests that the modulation process fails to reproduce lower frequencies accurately, making the voice sound unnaturally high and lacking depth.

Maximum Pitch Reduction: Robotic voices have a lower maximum pitch (256.64 Hz) compared to human voices (275.77 Hz), indicating a reduction of 19.12 Hz. This compression of the dynamic range makes the voice sound less expressive and more monotonous, posing challenges for forensic analysis.

Reduced Number of Pulses: Robotic voices show fewer voiced sound periods (560.81) compared to human voices (888.47), a decrease of 327.66 pulses. This affects the natural rhythm, making the speech sound less fluid and more segmented due to simplified speech patterns during modulation.

Altered Number of Periods: The number of periods in robotic voices (507.45) is reduced compared to human voices (859.97), a difference of 352.52 periods. This alteration impacts perceived continuity, making the voice sound less natural and more artificial.

Increased Period Variability: Robotic voices show slight irregularities in the timing of speech periods, with a standard deviation of 0.968 compared to 0.923 in human voices, an increase of 0.045. This adds to the mechanical quality, distinguishing them from natural human speech.

Decreased Voice Breaks: Robotic voices have fewer interruptions (16.3) compared to human voices (25.43), a reduction of 9.13 breaks. This results in smoother but less naturally interrupted speech, making the voice sound less human.

Higher Fraction of Unvoiced Frames: Robotic voices contain a higher proportion of unvoiced frames (61.69%) compared to human voices (44.20%), an increase of 17.49%. This contributes to a mechanical sound, as natural speech typically has fewer unvoiced segments.

Increased Jitter: Robotic voices exhibit greater frequency instability with a jitter value of 289.04 compared to 144.76 in human voices, an increase of 144.28. This makes the voice sound less smooth and more artificial.

Higher Autocorrelation: Robotic voices show higher autocorrelation values (1.426) compared to human voices (0.929), an increase of 0.497. This indicates more repetitive patterns, making the voice sound more robotic and less natural.

Increased Noise-to-Harmonics Ratio: Robotic voices have a higher noise-to-harmonics ratio (0.3627) compared to human voices (0.0875), an increase of 0.2752. This reduces clarity and introduces distortion, making the voice sound less natural and clear.

Decreased Harmonics-to-Noise Ratio: Robotic voices have fewer harmonic components relative to noise with a harmonics-to-noise ratio of 5.24 compared to 14.63 in human voices, a reduction of 9.39. This contributes to a more distorted and less clear sound, reducing perceived naturalness and clarity.

CONCLUSION

The comprehensive analysis using PRAAT provides insights into the differences between human and AI-modulated robotic voices, revealing significant alterations in various acoustic parameters due to AI modulation, with implications for forensic voice identification and authentication. The study found that median and mean pitch values are significantly higher in robotic voices, indicating a shift towards higher frequencies and affecting the natural pitch characteristics of the original speaker. The standard deviation of pitch is slightly higher in robotic voices, contributing to a more mechanical sound. The compression of pitch range reduces the dynamic range of the voice, resulting in less expressiveness. The reduction in the number of pulses and periods in robotic voices reflects a simplified speech pattern,

and the slight increase in the standard deviation of period suggests more irregularity in speech timing. The reduction in voice breaks indicates a smoother but less naturally interrupted speech pattern, while the increase in locally unvoiced frames contributes to the mechanical nature of the voice. The substantial increase in jitter indicates greater frequency instability, making the voice sound less smooth and more artificial. Higher mean autocorrelation values suggest more repetitive patterns, and the increased noise-to-harmonics ratio reflects more noise, resulting in a more distorted voice. These findings demonstrate that AI modulation significantly alters various acoustic parameters, making the robotic voice sound distinctly different from the human voice. PRAAT's analysis can reliably identify these differences, but identifying the original speaker solely based on these parameters is challenging due to the significant alterations introduced by AI modulation. However, the distinct differences provide a strong basis for excluding AI-modulated voices from being identified as the original human voice in forensic analysis, highlighting PRAAT's effectiveness in differentiating between natural and artificial voices for forensic applications.

LIMITATIONS

- Limited Scope of Vocal Variability: While the study explores the effects of gender and age on speaker identification, vocal characteristics can be further influenced by ethnicity, language background, and even health conditions. Including these additional variables in future studies could lead to more comprehensive and nuanced results.
- Restricted Speech Sample: The analysis employed a single threatening script, which limits the natural variation present in spoken language. Different emotions, speaking styles, and content can all influence how AI modulation affects speaker identification. Future studies should incorporate a wider range of speech samples to achieve a more generalizable understanding.
- Limited Existing Research: The current study lacks extensive supporting research or established publications. This creates a gap in terms of benchmarks and existing frameworks to guide further investigation. Additional research efforts are necessary to establish a robust foundation for future studies in this area.

REFERENCES

[1] Anand, S., Kopf, L. M., Shrivastav, R., & Eddins, D. A. (2021). Using pitch height and pitch strength to characterize type 1, 2, and 3 voice signals. *Journal of Voice, 35*(2), 181-193. https://doi.org/10.1016/j.jvoice.2019.08.006

[2] Baskoro, A. B., Cahyani, N., & Putrada, A. G. (2020). Analysis of Voice Changes in Anti Forensic Activities: Case Study of Voice Changer with Telephone Effect. *International Journal on Information and Communication Technology (IJoICT, 6*(2), 64-77. https://doi.org/10.21108/ijoict.v6i2.508

[3] Chen, W., & Jiang, X. (2023). Voice-Cloning Artificial-Intelligence Speakers Can Also Mimic Human-Specific Vocal Expression. *Preprints*, 2023120807. https://doi.org/10.20944/preprints202312.0807.v1

[4] Grillo, E. U., & Wolfberg, J. (2023). An Assessment of Different Praat Versions for Acoustic Measures Analyzed Automatically by VoiceEvalU8 and Manually by Two Raters. *Journal of Voice, 37*(1), 17-25. https://doi.org/10.1016/j.jvoice.2020.12.003

[5] Hiremath, B. N., & Patil, M. M. (2019). Analysis of speech in human communication. *J.S.S Academy of Technical Education, Bengaluru, Karnataka, India*. https://doi.org/10.5281/zenodo.3250518

[6] Huang, K. L., Duan, S. F., & Lyu, X. (2021). Affective Voice Interaction and Artificial Intelligence: A research study on the acoustic features of gender and the emotional states of the PAD model. *Frontiers in Psychology, 12*, 664925. https://doi.org/10.3389/fpsyg.2021.664925

[7] Hughes, V., Foulkes, P., & Wood, S. (2016). Strength of forensic voice comparison evidence from the acoustics of filled pauses. *International Journal of Speech, Language and the Law, 23*(1), 99-132. https://doi.org/10.1558/ijsll.v23i1.29874

[8] Irfan, M., Ramdania, D. R., Hasni, N., Budiman, I., Maylawati, D. S., & Manaf, K.

(2021). Similarity Level Analysis of the Voices of Twins Using the Analysis of Variance and Likelihood Ratio Methods. In *2021 7th International Conference on Wireless and Telematics (ICWT)* (pp. 1-5). Bandung, Indonesia. https://doi.org/10.1109/ICWT52862.2021.9678443

[9] Karakoç, M. M., & Varol, A. (2017). Visual and auditory analysis methods for speaker recognition in digital forensic. In *2017 International Conference on Computer Science and Engineering (UBMK)* (pp. 1113-1116). Antalya, Turkey. https://doi.org/10.1109/UBMK.2017.8093505

[10] Laia, Y. E. (2021). The Comparison Analysis of Voice Level Through Pitch and Formant Value Identification Techniques Through Praat Software. IJFL (International Journal of Forensic Linguistic), 2(2), 90-97. https://doi.org/10.22225/ijfl.3.1.4614.90-97

[11] Lovato, A., De Colle, W., Giacomelli, L., Piacente, A., Righetto, L., Marioni, G., & de Filippis, C. (2016). Multi-Dimensional Voice Program (MDVP) vs Praat for Assessing Euphonic Subjects: A Preliminary Study on the Gender-discriminating Power of Acoustic Analysis Software. *Journal of Voice, 30*(6), 765.e1-765.e5. https://doi.org/10.1016/j.jvoice.2015.10.012

[12] Magdin, M., Sulka, T., Tomanová, J., & Vozár, M. (2019). Voice analysis using PRAAT software and classification of user emotional state. *International Journal of Interactive Multimedia and Artificial Intelligence, 5*(6), 33-42. https://doi.org/10.9781/ijimai.2019.03.004

[13] Sampaio, M. C., Bohlender, J. E., & Brockmann-Bauser, M. (2021). Fundamental frequency and intensity effects on cepstral measures in vowels from connected speech of speakers with voice disorders. *Journal of Voice, 35*(3), 422-431. https://doi.org/10.1016/j.jvoice.2019.11.014

[14] Singh, S., Kumar, S., & Ali, A. (2021). Study on Variation in Speaker Identification under Different Conditions. *Indian Journal of Forensic Medicine and Pathology, 14*(2). DOI: http://dx.doi.org/10.21088/ijfmp.0974.3383.14221.44

[15] Sondhi, S., Khan, M., Vijay, R., Salhan, A. K., & Chouhan, S. (2015). Acoustic analysis of speech under stress. *International Journal of Bioinformatics Research and Applications, 11*(5), 417-432. https://doi.org/10.1504/IJBRA.2015.071942

[16] Sondhi, S., Vijay, R., Khan, M., & Salhan, A. K. (2016). Voice analysis for detection of deception. In 2016 11th International Conference on Knowledge, Information and Creativity Support Systems (KICSS) (pp. 1-6). Yogyakarta, Indonesia. https://doi.org/10.1109/KICSS.2016.7951455

[17] Surahman, A. (2021). An analysis of voice spectrum characteristics to the male voices recording using praat software. *IJFL (International Journal of Forensic Linguistic)*, *2*(2), 69-74. https://doi.org/10.22225/ijfl.3.1.4776.69-74

[18] Vestman, V., Kinnunen, T., González Hautamäki, R., & Sahidullah, M. (2019). Voice mimicry attacks assisted by automatic speaker verification. https://doi.org/10.48550/arXiv.1906.01454

[19] Wirdayanthi, A. I. (2021). The utilization of Praat program in determining the authenticity of the voice. *International Journal of Forensic Linguistics (IJFL)*, *2*(2), 81–89. https://doi.org/10.22225/ijfl.3.1.4774.81-89

[20] Xie, Z., Gadepalli, C., Farideh, J., Cheetham, B. M. G., & Homer, J. J. (2018). Machine Learning Applied to GRBAS Voice Quality Assessment. *Advances in Science, Technology and Engineering Systems Journal, 3*(6), 329-338. ISSN 2415-6698. Available at http://clok.uclan.ac.uk/25743/