Bridging Audio Forensics and Forensic Psychology for comparative analysis of natural and mimicked speech

P.S.Marathe[1], D.P.Madhusudan[2], Kaashika[3], Balasaheb.J.Nagare[4]

[1] PhD Student, University Department of Physics, University of Mumbai and Scientific Officer, RFSL, Pune

[2] Scientist B, Forensic Psychology, CFSL, Pune, MHA, India

[3] Forensic Professional, Forensic Psychology, CFSL, Pune, MHA, India

[4] Head of the Department, Professor at University Department of Physics, University of Mumbai

Abstract: This study examines the integration of audio forensics and forensic psychology in detecting similarities between natural and mimicked voices. Ten mimicry artists were instructed to replicate a predetermined script in target voices, producing both natural and mimicked recordings. Traditional audio forensic techniques, focusing on phonetic and acoustic parameters, were applied to assess the degree of similarity, while Layered Voice Analysis (LVA) was employed to identify deception-related indicators such as emotional stress, cognitive load, and psychophysiological cues. Results revealed that formant frequency analysis distinguished clear differences between natural and mimicked speech, whereas auditory evaluation highlighted notable similarities within each artist's samples. LVA further confirmed these similarity markers, demonstrating its potential utility in mimicry-related cases. The findings emphasize a multidisciplinary approach, where combining acoustic evidence with psychological indicators enhances the reliability of forensic voice analysis. This integrative framework strengthens evidentiary assessments in criminal investigations and supports the judicial interpretation of voice evidence.

Keywords: *Audio forensics, Layered Voice Analysis, Deception Detection, Forensic Psychology, voice mimicry.*

1. INTRODUCTION

The most basic definition of *Forensics* originates from the Latin word *'forensis'* meaning *public, to the forum or public discussion*; while the most modern definition of *forensics* would amount to *used in, or suitable to a court of law* [1][2]. Thus, it can be rightly said that any field of science that is used for legal purposes can be considered to be part of the field of forensic sciences; and can be considered vital tools in any legal proceedings be it civil or criminal [3][4].

Recent trends in forensic sciences have emerged with a multi - disciplinary approach, as compared to the initial days of trying to understand the basic usage of forensic sciences within the areas of criminal investigations. The evolving trends of forensic sciences in conjunction with the technological advances show there is an increasing need for a comprehensive and accurate analysis

of evidences involved in any legal dispute. It can, therefore, be rightly said that forensic sciences is not limited to the traditional laboratory based disciplines like biological or chemical sciences but has now integrated newer fields like cyber forensics, behavioral forensics and/or data sciences.

With multiple technological advancements and innovations such as next-generation sequencing in DNA analysis, excellence centers in digital forensics, and AI-powered tools, forensic science is undergoing revolution [5]. So when we talk about collaboration amongst different forensic disciplines, it is because it allows for cross-verification of findings, reducing the likelihood of errors and misinterpretations, and ensuring the reliability of forensic evidence [6].

With the increasing complexity of crimes along with the advancements in technology, this has created a desperate need for the integration of diverse forensic science techniques in criminal investigations. This integration, therefore would allow for a far more comprehensive analysis of evidences, thus leading to an accurate and sound investigations with hopefully an improved outcomes during the trials [5], [7].

2. FORENSIC AUDIO ANALYSIS AND LAYERED VOICE ANALYSIS EXAMINATION: A RESEARCHER'S PERSPECTIVE

Two such possible areas of integration within the fields of forensic sciences are the audio analysis and behavioral sciences; specifically Layered Voice Analysis Examination. Both the areas use audio recordings; wherein the audio analysis of the audio is conducted to analyze the auditory and spectral similarities of the speaker's voices, whereas the Layered Voice Analysis of the audio recordings is used to determine the deception patterns in the emotional, stress and cognitive aspects within the speaking patterns of the individual while narrating the event under investigation.

Most recent advancements in audio forensic methodologies have provided empirical evidence of using acoustic - phonetic analysis for both, speaker identification/verification as well as mimicry detection. Research showed that the pitch contour, duration and amplitude, more commonly known as Prosodic features tend to be favored over other spectral features because of their interpretability and robustness that tend to vary across other external conditions such as background noise as well as channel mismatch [8]. The threat posed by mimicry or in other words, voice spoofing has found its niche within the empirical research by application of prosodic and formant analyses that can be used to identify imposter speech.

Parallelly, Layered Voice Analysis is a recent technological software that is used to detect psychological indicators of deception, such as stress, cognitive load and emotional cues within the voice recording of the individual. Early empirical evaluations have demonstrated that there are chance-level performance (42–56% true-positive rates) and high false-positive rates (40–65%) when LVA was assessed under controlled laboratory conditions [9][10]. There were other similar

skepticism echoed in reviews that have challenged the scientific validity of voice-stress technologies, cautioning that LVA may lack sensitivity and robustness without strong operator influence [11]. Recent research however, also reinforces this viewpoint, stating that the human analysis or interpretation often outperformed the automated LVA judgments (68–71% accuracy for auditors versus 48% for automated systems); thus, suggesting that operator bias and experience could play significant roles in the analysis and interpretation of the LVA results [12].

Voice mimicry has been investigated in automatic speaker verification systems concluding that while untrained impersonators could produce prosodic features but could not produce more than simple modest alterations in formant and spectral features which indicates limited threat from simple mimicry [13]. However, it is pertinent to mention that high; lying trained or rather, expert mimicry is still an under explored challenged within the forensic audio analysis and overall forensic context.

This analysis of the literature findings clearly underscore the necessity of having an integrated and complimentary forensic analysis of evidences such as mimicked voices. Here, Acoustic - phonetic as well as phonological analysis can determine the similarities of the recorded voices. On the other hand, Layered Voice Analysis (LVA) can offer psychological insight into the cognitive stressors within the recorded voice - especially with a forensic psychologist interpreting the software results.

3. METHODOLOGY

The present research has adopted an experimental within-subjects design with the aim to investigate the forensic implications of mimicked voice recordings using complimentary techniques of two different fields. This research has focused on evaluating the authenticity of the voice recording along with the deception patterns in the voice samples through a dual-method approach: acoustic-phonetic analysis for speaker verification/identification and Layered Voice Analysis (LVA) for deception detection.

The design includes controlled recording sessions of mimicry artists producing both their natural voice and mimicked versions using a fixed script, enabling comparison across both forensic modalities.

3.1. Participants

For the current research, the authors invited Ten trained mimicry artists (n=10 named as A to J), each with demonstrated proficiency in vocal imitation. Participants were recruited through purposive and snowball sampling technique through interconnected network of mimicry artists. Participants were duly informed regarding the study's aim and the procedures involved within the study and only after they provided informed consent, the voice recording were taken.

3.2. Materials

Each participant was given a standardized script of 150–200 words that was designed to elicit a range of prosodic and semantic features. Recordings were conducted in a contained environment. The mimicry task was recorded in two phases: (1) the participant's natural voice (A1, B1, C1…), and (2) their mimicked voice (A2, B2, C2...) of a designated public figure with a pre - designed script in the regional language. Audio files were processed using KAY PENTAX's Multispeech and Praat softwares for acoustic-phonetic analysis and Nemesysco's Layered Voice Analysis (LVA-i) software for deception metrics.

3.3. Procedure

Each participant performed two recording sessions. In Session 1, they read the script in their natural speaking voice. In Session 2, they delivered the same script while mimicking the voice of a selected target speaker. Each session was recorded separately under consistent audio settings. To minimize performance variability, participants were allowed a short rehearsal period. All samples were anonymized using a randomized coding system prior to analysis to reduce analyst bias.

4. Audio Analysis

A growing body of forensic-phonetic work shows that "mimicked" or impersonated voices can fool casual listeners, yet close auditory-perceptual analysis often exposes patterned deviations from a target speaker. Classic imitation studies argue that while mimics can shift global prosody (overall pitch, speaking rate, rhythm), fine-grained articulatory timing and segmental habits remain stubbornly speaker-specific and are hard to copy consistently, giving trained listeners cues in stress placement, co articulation, and micro-rhythm [14]

Researchers investigating mimicked (impersonated) speech have converged on a set of auditory parameters that repeatedly appear in both production (acoustic) and perceptual studies. The literature groups these parameters into prosodic features, voice-quality / phonation features, spectral / vocal-tract features, perturbation and harmonicity measures, and temporal / segmental measures — and shows that both what mimics try to copy and what listeners use to judge similarity are drawn from these domains.

a) Prosodic features: Fundamental frequency (F0) — its mean, dynamic range and the shape of intonation contours — is one of the most-studied dimensions because mimics often target a speaker's

pitch pattern and melodic shape. Studies show mimics can approximate global prosody (overall pitch level and contour) better than fine-grained timing, and listeners rely heavily on prosodic similarity in perceptual tests [15].

b) Voice-quality / phonation features: Imitators commonly alter phonation type to match a target (e.g., creaky voice, falsetto, breathy voice). Case studies and systematic analyses report that such phonation changes often leave tell-tale artifacts (unstable vibration, strain) detectable by auditors and by spectral inspection; some disguises (e.g., deliberate glottal fry) have been specifically investigated for their effectiveness and audibility [16, 17].

c) Spectral / vocal-tract features: Vocal-tract resonances (formant patterns F1–F3), long-term average spectrum (LTAS) and spectral tilt are strong correlates of speaker identity and are therefore key targets for mimics. Empirical work on professional impersonators shows partial success in shifting formant locations toward targets, but systematic differences often remain and are measurable with formant analysis and LTAS comparisons [18, 19].

d) Perturbation and harmonicity measures: 4. Measures of cycle-to-cycle stability — jitter, shimmer, and harmonic-to-noise ratio — capture micro variations in vocal fold vibration and are sensitive to strained or deliberately altered phonation. Several studies recommend these as objective indices that reveal artifacts of sustained disguise attempts [20, 21].

e) Temporal / segmental measures: Rhythm, speech rate, placement of pauses, and fine segmental timing (including co-articulatory patterns) are robustly speaker-specific and comparatively resistant to accurate imitation. Research shows that even skilled impersonators often leave native timing and co-articulatory “habits” of their own speech, which trained listeners—and quantitative temporal analyses—can exploit [22, 23].

In summary, the literature shows that mimics tend to succeed more on coarse prosodic and perceptual dimensions than on fine-grained spectral, perturbation and temporal habits that encode much of speaker identity. Forensic and experimental researchers therefore treat a combined approach of auditory parameters (prosody, phonation, spectral, perturbation, timing) as the standard approach to studying and distinguishing mimicked from authentic speech [18,20].

4.1 Observations-Audio analysis

In the analysis of mimicked and original speech, it is often observed that mimicry artists skillfully reproduce temporal aspects of the target speaker's delivery. They can imitate the same timing, rhythm, and placement of pauses, which creates a strong impression of similarity to the casual listener. This ability to mirror prosodic patterns—such as speaking rate and pause structure—enhances the perceptual plausibility of the imitation.

In the present analysis, it was observed that the mimicry artist employed the same pauses and speech rate while producing both the mimicked speech and his own natural voice. This finding suggests that temporal features such as rhythm, pacing, and pause placement are part of the artist's inherent speaking style and remain consistent across both speech conditions. Although these prosodic similarities may enhance the perceptual credibility of the imitation, they also indicate that temporal patterns are less speaker-specific and can be influenced by the habitual style of the mimic rather than the target speaker's unique identity. Speech rate is often an ingrained characteristics of an individual's natural speaking pattern, which the mimicry artist unconsciously carried into both natural and mimicked speech. However, when the spectral characteristics of the speech are examined, clear differences emerge. This contrast highlights why auditory analysis, supported by acoustic examination, is essential in distinguishing genuine speech from mimicry.

In current study, the jitter and shimmer values of the mimicry artist's mimicked speech were found to be approximately the same as those observed in his natural speech. Jitter (cycle-to-cycle variation in fundamental frequency) and shimmer (cycle-to-cycle variation in amplitude) are measures that reflect the stability of vocal fold vibration. The similarity in these values across both speaking conditions suggests that the phonatory mechanism of the artist remained consistent, regardless of whether he was producing his natural or mimicked voice. This indicates that, although the artist modified certain suprasegmental or spectral features for imitation, the underlying vocal fold physiology and phonatory stability were not significantly altered. Hence, jitter and shimmer serve as reliable indicators of the speaker's own vocal identity, even when attempts at mimicry are made.

It is generally observed that a mimicry artist tends to choose and imitate voices whose pitch range is close to his or her own natural pitch. Since pitch is primarily determined by the physiological characteristics of the vocal folds (such as length, tension, and mass), large deviations from the artist's habitual pitch are difficult to sustain without vocal strain or instability. By selecting target voices with a similar fundamental frequency (F0) range, the artist can maintain natural phonatory stability

while focusing on other features such as intonation, pauses, and voice quality to enhance the illusion of similarity. This strategy allows the mimicry to sound more convincing to listeners while avoiding vocal fatigue.

Another observation is stress, the stress patterns on particular consonant clusters in the mimicry artist's mimicked speech were found to be approximately the same as those in his natural speech. Stress refers to the relative prominence given to certain syllables or clusters through variations in intensity, pitch, and duration. The similarity of stress placement across both speaking conditions indicates that the mimicry artist unconsciously retained his habitual prosodic framework, even while attempting to imitate another voice. This suggests that stress distribution is a stable, speaker-specific feature, rooted in the individual's articulatory habits and motor-speech control. While the artist may modify pitch, timbre, or voice quality to enhance the illusion of mimicry, the consistent use of stress on clusters reflects the persistence of underlying speech patterns that can serve as reliable markers for forensic voice comparison and speaker identification. In Fig. 1 first column waveform and spectrograph of mimicked word and in the second column is that of the word from natural speech. The stressed syllables are represented as the darker and well defined formant patterns. Here all the consonants /b/, /b/ /s/ in word "Babasaheb" are equally stressed.

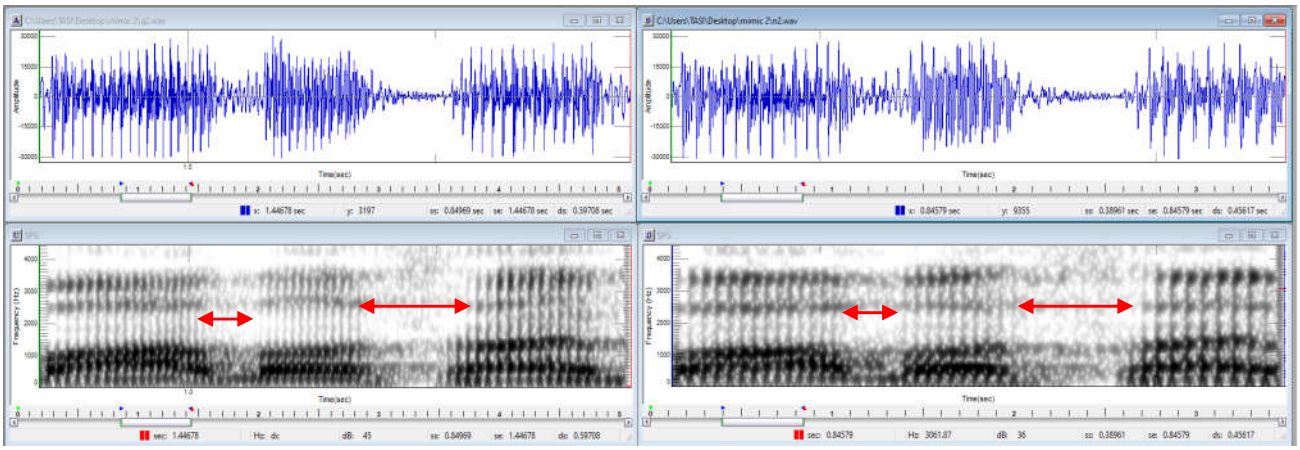


Figure 1: Spectrographic representation of Word "**Babasaheb**" in mimicked and Natural Speech

Auditory analysis plays a critical role in forensic investigations involving mimicry because it enables experts to detect subtle differences between an impersonated voice and the original speaker. While a mimicry artist may successfully reproduce speech rate, pauses, rhythm, and certain prosodic patterns, fundamental spectral characteristics, vocal quality, micro-prosody, and phonatory features often remain distinct. By carefully examining intonation, stress, timing, formant patterns, jitter, shimmer,

and other voice-quality markers, forensic analysts can identify inconsistencies that are not perceptible to untrained listeners.

5. Layered Voice Analysis

Voice recordings were also analyzed using Layered Voice Analysis (LVA-e) software, which claims to detect cognitive, emotional, and stress-related deception markers in voice recordings of the individual. The LVA assessment focused on five primary deception-related indicators: emotions, stress, cognition, thinking and anticipation levels. Graphical analysis of the voice recordings were conducted to see if deceptive stress patterns emerged more prominently during mimicked speech when compared to natural speech in the following parameters:

a) Global Risk Parameter: This parameter summarizes deviation across multiple indicators: emotional, cognitive and anticipatory indicators of deception in the voice modulations. Therefore, this is an estimation of the overall probability of the psychological and behavioral risk indicators of deception based on an integrated voice metrics [24]. This parameter mainly functions as a screening parameter that reflects the general instability or the tension in the speaker's state.

b) Emotional Stress: This parameter in LVA tries to analyze the emotional arousal or strain detected in the speaker's voice that remains independent of the semantic content, i.e. what the speaker says. The assumption here is that this affective strain is a reflection of psychophysiological activation (i.e. anxiety, stress and fear) that can be due to the narration of the version of events, especially when using in the forensic and investigations; for elevated emotional stress could correspond to heightened affective load during the narration [9] [24]

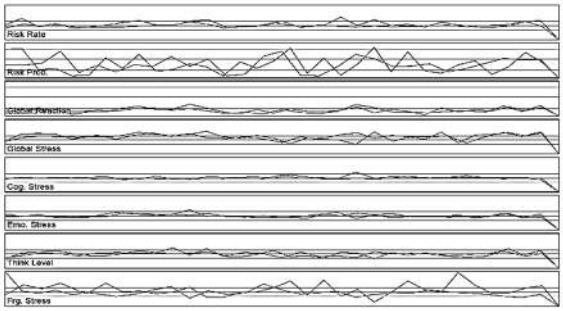
c) Cognitive Stress: This parameter, also known as Cognitive Load analyses the mental effort as well as the cognitive dissonance that the speaker may experience when narrating the event. Therefore, this parameter captures the signs of mental conflict, or suppression and discriminates it from the purely emotional strain as experienced by the speaker. In forensic terms, a higher cognitive stress level could indicate deliberate reasoning or even concealment of information [9].

d) Thinking Level: This parameter in LVA reflects the level of conscious reasoning, planning or even logical structuring that the speaker maybe experiencing within their speech, Therefore, this parameter indicates the level of speaker's engagement within the analytical and/or strategic thought processes when compared to spontaneous or emotionally driven responses [25].

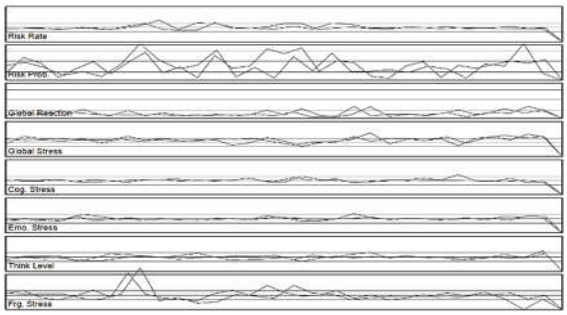
e) Anticipation Level: This parameter analyses how forward looking mental capacity of the speaker; i.e. the level of preparation the speaker has for future interactional turns w.r.t. the events he is narrating. This parameter, therefore, measures the psychological readiness of the speaker; that may manifest with subtle changes in rhythm, modulation or even the timing [25]. Within the forensic interview settings, elevated anticipation levels have been linked to pre - planned and/or strategic behavior.

5.1. Observations-Layered Voice Analysis

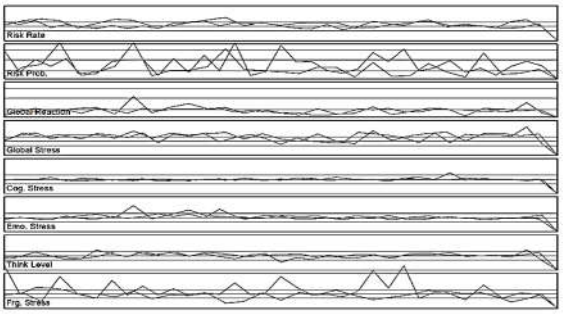
The ability to replicate a well-known celebrity's voice is a skill that mimicry artists take great pride in and have built their entire career upon. Therefore, it is not difficult to comprehend the emotional bond that the artist will have with this talent. The parametric comparison of the natural speech and the mimicked speech of the mimicry artists are given in the figure 2.



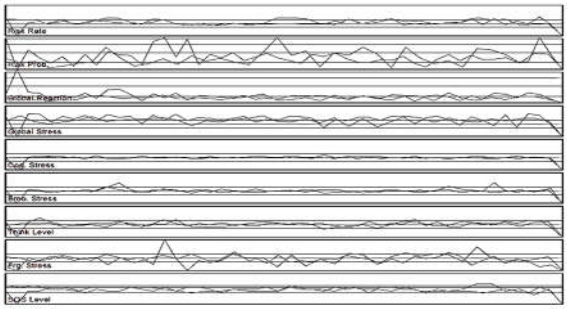
A1-A2



B1-B2



C1-C2



D1-D2

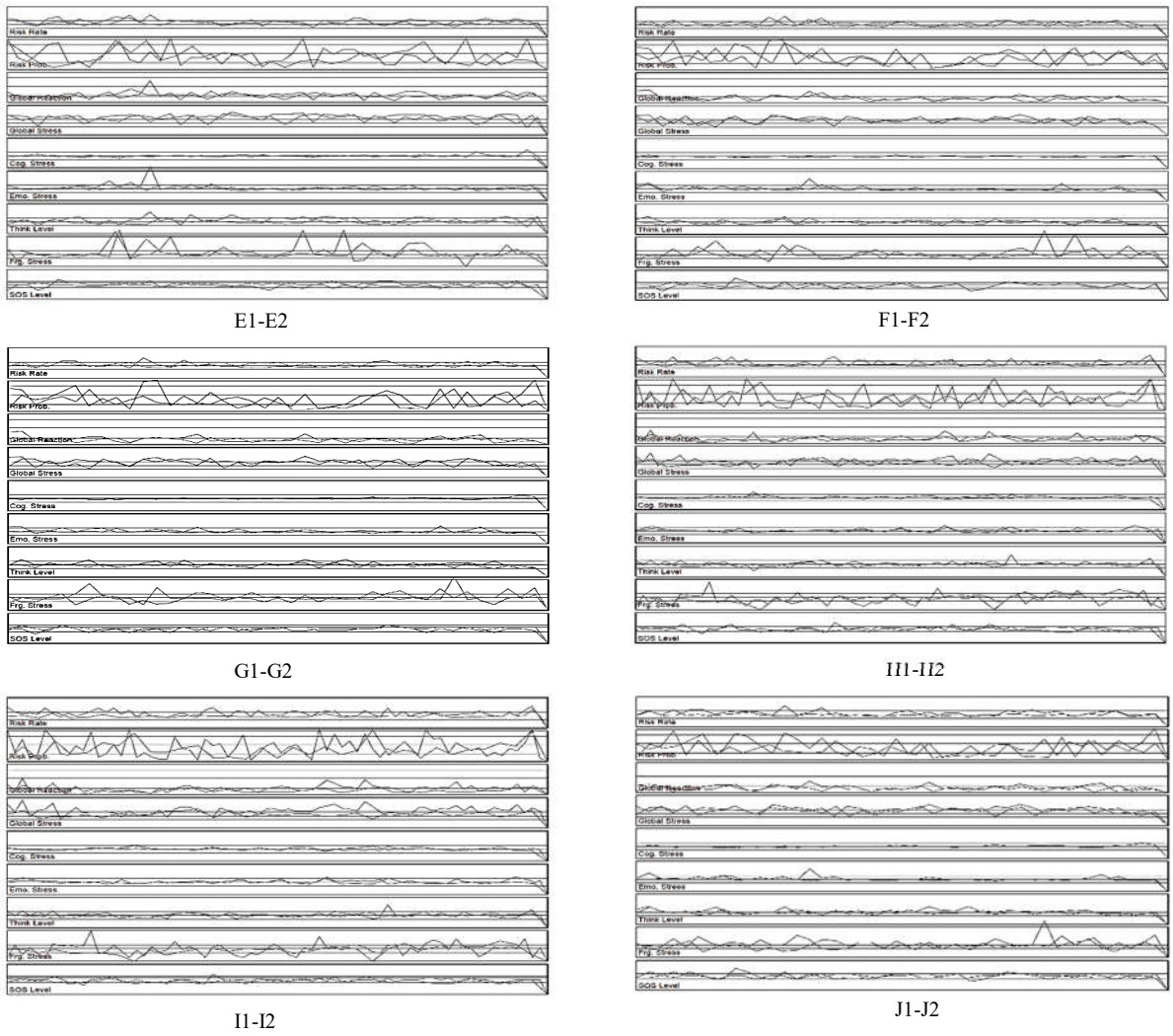


Figure 2: The parametric comparison of the natural speech and the mimicked speech of the mimicry artists

The above figure 2 clearly shows that there are similarities in the basic five parameters of emotional stressors on Layered Voice Analysis, i.e. Global Risk, Emotional Stress, Cognitive Stress, Thinking Level and Anticipation Level across all the twenty different recordings. Both natural and mimicked speech carry traces of the artist's own emotional baseline. Even when attempting to replicate another voice, subtle markers such as micro-fluctuations in pitch or vocal intensity can reveal the artist's personal emotional state. Thus, emotions act as a common underlying layer in both voice types. Whether producing natural or mimicked speech, the artist relies on heightened thinking processes such as attention, planning, and linguistic structuring. Both forms of speech involve active monitoring of pronunciation, prosody, and articulation, indicating a shared cognitive mechanism at the thinking level. In both cases, the artist experiences performance-related stress. The natural voice may reflect stress in authentic form, whereas the mimicked voice reflects controlled stress due to the demand for accuracy. However, the physiological stress response (e.g., vocal tremor, breath control variations) is common to both.

Memory, auditory discrimination, and motor control are equally engaged in both natural and mimicked speech. For natural speech, these processes support fluent expression; for mimicked speech, they are redirected toward adapting speech to the target voice. Nonetheless, the core cognitive resources remain the same. The natural and mimicked voices of a mimicry artist share similarities in emotional undercurrents, cognitive processing, stress response, and thinking patterns. These commonalities explain why mimicked speech may perceptually resemble the natural voice of the artist while still bearing distinct acoustic and psychological markers for forensic detection.

6. Discussion

The study's inferences shed crucial light on the intricate forensic issues surrounding mimicking in audio recordings. Conventional audio forensic methods, like acoustic and phonetic studies, have proven effective in identifying structural variations between the genuine and imitated voices, especially in formant frequency patterns. A high degree of perceptual similarity was also discovered via auditory analysis, highlighting the fact that mimicry can pose a serious risk to the reliability of voice testimony in court. The constraints of using only auditory-based or acoustic characteristics for evidentiary purposes are highlighted by this duality. The integration of forensic psychology through Layered Voice Analysis (LVA) offered a critical complementary perspective. Unlike traditional forensic methods, LVA was able to identify deception-related markers, including indicators of emotional stress and cognitive load, which are difficult to mask even in well-executed mimicry. These findings suggest that psychological markers provide an additional layer of robustness in authenticity assessment, bridging gaps left by purely acoustic or perceptual techniques. A key implication of this research is the value of a multidisciplinary approach. By combining audio forensic science with forensic psychology, investigators are better positioned to differentiate between genuine and manipulated voice recordings. Such integration is particularly relevant in the courtroom, where the admissibility of voice evidence often depends not only on scientific validity but also on the interpretability of the findings by legal professionals.

Nevertheless, certain limitations must be acknowledged. The structured experimental design involved mimicry artists replicating predetermined scripts, which may not capture the full variability of spontaneous speech in real-world criminal contexts. Additionally, LVA, while promising, is not without criticism regarding its reliability and validity in different linguistic and cultural settings. Future research should expand on this work by testing larger and more diverse datasets, incorporating spontaneous speech, and validating LVA against other psychophysiological and computational deception-detection tools. Overall, the study advances the understanding of how

mimicry challenges the evidentiary value of voice recordings and underscores the necessity of integrating technological and psychological perspectives in forensic practice. Such integration has the potential to refine the standards of voice authentication, ultimately aiding judicial systems in making more reliable determinations when voice evidence is presented.

7. Conclusion

This study demonstrates that mimicry poses a serious challenge to the authenticity and admissibility of voice evidence, as auditory analysis often highlights perceptual similarities despite underlying acoustic differences. By integrating audio forensic methods with forensic psychological tools such as Layered Voice Analysis, it becomes possible to detect deception markers that remain hidden in traditional analyses. The findings underscore the importance of a multidisciplinary approach to strengthen the reliability of voice evidence in legal proceedings. Future research should further validate this integrative framework across diverse speech contexts to enhance its applicability in forensic practice. Future investigations should extend this work by involving a larger pool of participants to improve the representativeness and statistical validity of the findings. With the increasing prevalence of AI-generated and deepfake voices, research must also adapt to assess how integrative forensic approaches can effectively distinguish between natural, mimicked, and synthetic speech. Furthermore, developing text-independent and language-independent analytical frameworks will be essential for ensuring that these methods remain robust and universally applicable across diverse linguistic and cultural contexts. Such advancements will enhance the reliability, adaptability, and global relevance of forensic voice analysis in both investigative and judicial domains

8. Acknowledgement

Authors would like to thank Mr. Sanjay Kumar Verma, IPS, Director General, Legal and Technical, Directorate of Forensic Science Laboratories, Mumbai, India for unwavering support and Dr. V. J. Thakare Director, Directorate of Forensic Science Laboratories, Mumbai for providing research environment and guidance throughout the work. We sincerely thank Dr. Ravindra Sharma, Director, Centre Forensic Science Laboratory, Pune for providing access to the Laboratory facilities. Their support was instrumental in carrying out this research work. We also gratefully acknowledge all the participants for their valuable involvement in the study.

REFERENCES

[1] Katz, E. (2015). Forensic Science - Multidisciplinary Approach. Forensic, Legal & Investigative Sciences, 1(1), 1–3. https://doi.org/10.24966/flis-733x/100004

[2] http://www.merriam-webster.com/dictionary/forensic

[3] James SH, Nordby JJ, Bell S (Eds.) (2009) Forensic Science: An Introduction to Scientific and Investigative Techniques (3rdedn). CRC Press, Boca Raton, USA.

[4] Siegel JA and Mirakovits K (2010) Forensic Science: The Basics (2nd edn). CRC Press, Boca Raton, USA.

[5] Chango, X., Flor-Unda, O., Gil-Jiménez, P., & Gómez-Moreno, H. (2024). Technology in Forensic Sciences: Innovation and Precision. *Technologies*, *12*(8), 120. https://doi.org/10.3390/technologies12080120

[6] Vladimir L (2023) Forensic Analysis: A Multidisciplinary Approach and the Collaborative Frontier. J Forensic Toxicol Pharmacol 12:3.

[7] Mishra, K., & Singh, A. (2024). Bridging the Gap: Integrating forensic science and legal frameworks in criminal justice. *International Journal of Applied Research*, *10*(12), 141–148. https://doi.org/10.22271/allresearch.2024.v10.i12c.12224

[8] Mary, L., Babu, K. K. A., & Joseph, A. (2012). Analysis and detection of mimicked speech based on prosodic features. *International Journal of Speech Technology*, *15*(3), 407–417. https://doi.org/10.1007/s10772-012-9163-3

[9] Harnsberger JD, Hollien H, Martin CA, Hollien KA. Stress and deception in speech: evaluating layered voice analysis. J Forensic Sci. 2009 May;54(3):642-50. doi: 10.1111/j.1556-4029.2009.01026.x. PMID: 19432740.

[10] Hollien, H., Geison, L., & Hicks, J. W., Jr (1987). Voice stress evaluators and lie detection. *Journal of forensic sciences*, *32*(2), 405–418.

[11] Wikipedia contributors. (2024, January 7). *Voice stress analysis*. Wikipedia. https://en.wikipedia.org/wiki/Voice_stress_analysis

[12] Horvath, F., McCloughan, J., Weatherman, D., & Slowik, S. (2013). The accuracy of auditors' and layered voice Analysis (LVA) operators' judgments of truth and deception during police questioning. *Journal of forensic sciences*, *58*(2), 385–392. https://doi.org/10.1111/1556-4029.12066

[13] Vestman, V., Kinnunen, T., Hautamäki, R. G., & Sahidullah, M. (2019). Voice mimicry attacks assisted by automatic speaker verification. *Computer Speech & Language*, *59*, 36–54. https://doi.org/10.1016/j.csl.2019.05.005

[14] Article-Wretling, Pr & Eriksson, Anders. (1998). Is articulatory timing speaker specific? - Evidence from imitated voices.

[15] Mary, Leena & Babu, Anish & Joseph, Aju & George, Gibin. (2013). Evaluation of mimicked speech using prosodic features. Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on. 7189-7193. 10.1109/ICASSP.2013.6639058.

[16] A. Hirson, M. Duckworth,Glottal fry and voice disguise: a case study in forensic phonetics,

Journal of Biomedical Engineering,Volume 15, Issue 3,1993,Pages 193-200,ISSN 0141-5425.

[17] Kitamura, Tatsuya. (2008). Acoustic analysis of imitated voice produced by a professional impersonator. Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH. 813-816. 10.21437/Interspeech.2008-248.

[18] R. Singh, D. Gencaga and B. Raj, "Formant manipulations in voice disguise by mimicry," 2016 4th International Conference on Biometrics and Forensics (IWBF), Limassol, Cyprus, 2016, pp. 1-6, doi: 10.1109/IWBF.2016.7449675.

[19] López, S., Riera, P., Assaneo, M. et al. Vocal caricatures reveal signatures of speaker identity. Sci Rep 3, 3407 (2013). https://doi.org/10.1038/srep03407

[20] Wertzner HF, Schreiber S, Amaro L. Analysis of fundamental frequency, jitter, shimmer and vocal intensity in children with phonological disorders. Braz J Otorhinolaryngol. 2005 Sep-Oct;71(5):582-8. doi: 10.1016/s1808-8694(15)31261-1. Epub 2006 Mar 31. PMID: 16612518; PMCID: PMC9441971.

[21] Li G, Hou Q, Zhang C, Jiang Z, Gong S. Acoustic parameters for the evaluation of voice quality in patients with voice disorders. Ann Palliat Med. 2021 Jan;10(1):130-136. doi: 10.21037/apm-20-2102. Epub 2020 Dec 31. PMID: 33440977.

[22] Dellwo V, Leemann A, Kolly MJ. Rhythmic variability between speakers: articulatory, prosodic, and linguistic factors. J Acoust Soc Am. 2015 Mar;137(3):1513-28. doi: 10.1121/1.4906837. PMID: 25786962.

[23] Zetterholm, E. (1997). Impersonation: a phonetic case study of the imitation of a voice. Working Papers, Lund University, Dept. of Linguistics; Vol. 46. Pp.269-287.

[24] Nemesysco Ltd. (2024). *LVA Technology Overview.* Retrieved from https://www.nemesysco.com/lva-technology/

[25] Ado-Tech. (2023). About Layered Voice Analysis (LVA). B-Trust Documentation. Retrieved from https://docs.ado-tech.com/books/b-trust/page/about-layered-voice-analysis-lva-V1O