IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

MULTILINGUAL VOICE IDENTITY ANALYSIS USING AI IN SOUTHERN INDIAN LANGUAGES.

[1]B.N.D.S. Prasanna, [2]Nagasshree M N, [3]Name of 3rd Author
[1]MSc Student, [2]Assistant Professor, [3]Designation of 3rd Author
[1]Department of Forensic Science,
[1]Garden City University, Bangaluru, India

Abstract: Verifying speaker authenticity and identifying AI-generated voices has become crucial for forensic and cybersecurity applications due to the increasing complexity of voice-based AI technology. In order to ascertain if two audio recordings were made by the same person, the project "multilingual voice identity analysis using ai in southern Indian languages" attempts to create a voice comparison system that can analyses speech samples in Tamil, Telugu, Malayalam, and Kannada. The system makes use of a simple voice- matching algorithm that analyses acoustic properties like pitch, frequency, intensity, and speaking rate while processing two inputs: one in English and one in a native South Indian language. Native speakers provided two recordings one in their home tongue and one in English—for the collection of voice samples. A total of 100 speech samples, 25 from each language group, were examined. The findings revealed an overall 84% match accuracy, with the following language-based accuracy: Telugu (80%), Tamil (88%), Kannada (76%), and Malayalam (92%). Better speaker consistency and intelligibility across language shifts are suggested by the greater accuracy in Malayalam. To aid comprehend the system's performance, data visualization techniques such as pie charts and bar graphs were used to show the matched vs unmatched outcomes. The viability of employing basic AI-based voice comparison techniques for multilingual speaker verification is demonstrated by this work, particularly in forensic settings where linguistic variation is substantial. Additionally, the technique may be used to identify identity theft and voice cloning. Future research might involve extending the dataset for wider application across multiple languages and speaker differences as well as using deep learning models to increase accuracy.

Index Terms: Speaker identification, voice recognition, forensic phonetics, multilingual speech analysis, AI voice detection, cyber forensics

INTRODUCTION

The study of human voice dates back to ancient civilizations. As early as Ancient Greece, philosophers like Aristotle and Galen Studied the anatomy of the throat and larynx. A study of voice is called as Vocology, one of the most intricate and individualized forms of communication is the human voice. It is an essential tool for communication ideas, feelings, identity, and social engagement. Vocal cords, also called vocal folds, are found in the larynx, which is where voice is formed when air from the lungs is expelled through the trachea. The vocal cords vibrate to produce sound when air moves through the glottis, then area between them.

The first sound is unpolished and unclear. The vocal tract's resonating chambers, which include the mouth (oral cavity), nose (nasal cavity), and throt (pharynx), then shape and refine it. Lastly, language is made up of discrete speech sounds called phonemes, which are formed by the cooperation of the articulators, which include the tongue, teeth, lips and palate. From airflow to articulation, the entire process occurs in a split seconds and is controlled by both cerebral regulation and muscle coordination from the language and speech related parts of the brain, including Wernicke's and Broca's regions.

Phonology is the study of how speech sounds work in particular language, whereas phonetics is the study of the physical characteristics of speech sounds. The physiological, auditory, and cognitive processes involved in communication are examined in broader domains including voice science and speech science. Additionally, speech production is related to disciplines such as neurolinguistics, which examines how the brain shapes language. Gaining knowledge of these fields lays the groundwork for evaluating and deciphering human speech patterns. Variations in vocal cord length, tension, vocal tract forms, and acquired speaking patters give each person a distinctive voice that might be described as a vocal fingerprint.

Voice is an important tool in forensic science for identifying people and confirming their identities. Comparing speech samples to see if they are forms the same individual is known as forensic voice analysis. This method is applied in situation when a speech recording might be utilized as evidence, such as threats, ransom demands, or anonymous tips. A variety of vocal traits, including pitch, frequency, speech rhythms, pauses, accent, and intonation, are examined by expects.

Spectrophotograms, which show the frequency and amplitude of speech with the time, are frequently used to visualize these characteristics. Furthermore, voice is now being researched for its susceptibility to deepfakes and voice cloning due to the growth of digital communication and artificial intelligence. Given this, forensic investigation is also attempting to identify AI generated voiced, which are being utilized more frequently in cybercrimes, fraud, and impersonation.

Voice stress analysis, which looks for deceit in minute variations in voice frequency and pitch under stress, is another crucial component of forensic speech analysis. It is occasionally used to help lie detection during investigations while its validity is still up for question. Additionally, by using speech characteristics to identify a speaker's potential origin or background, language and accent profiling can help detectives identify potential suspects.

Cloned voices in several languages can be used by attackers to more convincingly mimic people in cybercrime scenarios. For instance, it is possible to clone a multilingual individual to perpetrate fraud both domestically and abroad. Multilingual speech analysis therefore turns as a crucial forensic technique. Forensic scientists are able to determine if a speaker is human or artificial by examining languages switching patterns, pronunciation changes, and emotional tones across languages. These methods are a component of the expanding fields of cyber forensic and forensic phonetics which use digital technologies to fight sophisticated voice-based scams

Examining whether a person retains a consistent vocal identity in two languages their original tongue and English despite natural speech fluctuations is the major goal of the study. This is accomplished by the creation and use of a unique software program that can analyse voice recordings made by the same person in both languages.

The program extracts and compared a variety of acoustic characteristics, such as fundamental frequency(pitch), formant frequencies, spectrum energy, speech tempo, and loudness, using sophisticated voice processing algorithms, when creating a digital profile of the speaker's voice, these characterises are essential.

While surface level features, such accent or fluency, might vary from one language to another underlying vocal qualities are more consistent and can be accurately quantified. A similarity score or classification result that shows whether two samples are from the same speaker is determined by the system's ability to interpret voice samples and perform speaker matching. It recognizes the underlying voiceprint that is specific to each individual while taking into consideration the normal variations that arise in cross linguistic conversation. The program determines that the speech samples most likely come from the same person if the similarity score rises beyond a certain level. In the field of forensic science, where speaker identification might be crucial in resolving crimes including ransom calls, cyber impersonation, phony audio recordings, and anonymous threats, this skill has significant ramification. Artificial intelligence powered voice cloning technology has presented cyber security and low enforcement with new difficulties in the modern day.

With just a few seconds of audio, attackers may now mimic a person's voice using AI models to produce lifelike speech that can be exploited for identity theft, frauds, or the creation of fake audio evidence. Determining if a suspected audio recording actually belongs to the accused person and differentiating between real and artificially manufactured voices is crucial in these situations. In order to meet this demand, this project develops a system that can verify speakers of different languages.

This is particularly helpful in multilingual cultures like India, where people may communicate in both regional and international languages.

Additionally, the experiment highlights how crucial it is to use multilingual speech behaviour analysis as a forensic technique. One may learn a lot about a person's identity from their prosodic characteristics, pronunciation patterns, and language transitions. Researchers can get a more complete speakers' profile by comparing speech in native and English circumstances. In cyber forensic investigations, when audio evidence is frequently scarce, altered, or extracted from multilingual sources like social media, phone records, or surveillance, this becomes very important.

Python is an interpretable, high-level programming language that is renowned for its ease of use, readability, and adaptability. It is appropriate for a variety of applications as it supports several programming paradigms, such as procedural, object-oriented, and functional programming. Data analysis, machine learning, image processing, automation, and web development are just a few of the tasks made easier by Python's extensive ecosystem of third-party packages and robust standard library. It is the perfect language for both novices and experts due to its dynamic typing and simple syntax.

In the field of forensic research, Python has become a vital tool owing to its capacity to automate and analyse complicated data effectively. It is frequently used in digital forensics for activities including retrieving lost files, extracting metadata, interpreting forensic artifacts, and examining logs. Memory forensics is supported by tools like Volatility, while forensic experts may access disk images and file systems using libraries like pytsk3 and dfvfs. With libraries like librosa, pydub, and praat- parse mouth, Python is also utilized in audio and video forensics, particularly in speaker recognition, speech analysis, and detecting altered recordings.

Additionally, Python plays a key role in network research and cyber forensics, allowing professionals to track IP addresses, identify abnormalities, and keep an eye on questionable activity with tools like socket and scapy. Python is used in forensic data science to enable AI-based pattern recognition, machine learning, and statistical modelling for fraud detection, criminal profiling, and predictive analysis. All things considered, Python is a strong friend in contemporary forensic investigations, bridging the divide between technology and criminal justice, thanks to its versatility and the availability of specialist forensic libraries.

Flask is a popular Python web framework for creating web apps that is both lightweight and adaptable. Because it does not by default come with built-in features like form validation or database abstraction layers, it is regarded as a microframework, allowing developers to select the tools and modules that best suit their project requirements. Because of its well-known simplicity, Flask is a great option for novices while also having the capacity to enable more experienced users to create intricate web apps. It offers fundamental functions including request processing, routing, and an integrated development server. Integration with the Jinja2 templating engine, which enables the display of dynamic content in HTML templates, is one of its essential features. Flask is very adaptable as it allows extensions to be used to add features like database support, authentication, and more. Flask is widely used in academic and industrial contexts for full-stack web applications, API development, and prototyping because of its easily comprehensible syntax.

A way of producing code or language structure that is simple enough for people to comprehend and interpret is known as comprehensible syntax. Programming code that is understandable, logically structured, and written in a way that makes its meaning obvious is guaranteed by a comprehensible syntax. Because it prevents misunderstandings and errors, this is particularly crucial in collaborative settings when several engineers work on the same codebase.
Because they place an emphasis on simplicity and clarity, languages like Python are commended for having an understandable syntax that makes writing and maintaining code simpler for both novice and expert programmers. More generally, by emphasizing readability and comprehension over complexity, comprehensible syntax reduces the learning curve, increases the effectiveness of debugging, and encourages best practices in software development.

A biometric technique called speaker recognition uses a person's vocal features to identify or confirm their identification. This procedure depends on each person's own vocal tract characteristics, speaking patterns, pitch, and rhythm. Speaker identification, which distinguishes the speaker from a collection of recognized speakers, and speaker verification, which verifies if the speaker is who they say they are, are the two main categories under which speaker recognition falls. The two primary stages of the system are usually enrolment and recognition. The speaker's voice is captured and processed during enrolment to provide a distinct voice print or model.

During the recognition phase, methods like Gaussian Mixture Models (GMM), Mel-Frequency Cepstral Coefficients (MFCC), or contemporary deep learning algorithms are used to compare a new speech sample with previously recorded voice prints. Speaker recognition is used in smart gadgets, forensic investigations, security systems, and customer service authentication. It is a developing area of artificial intelligence and speech processing as its efficacy is dependent on variables including background noise, microphone quality, speech unpredictability, and language

2. METHODOLOGY

AIM:

To create and test a speaker identification system that compares voice samples from the same person in different South Indian languages, to check if the voices match and identify any changes.

OBJECTIVES:

1. To examine and contrast the same person’s recordings of their native and English voices in order to find any variations or patterns in their vocal characteristics.

2. To create and put into use a software program that automatically assesses if two audio samples are from the same person.

3. To investigate how gender, regional phonetics traits, and bilingualism affect the accuracy of speaker identification.

Procedure of data collection:

1) The sample was collected from the university campus.
2) Collected from both male and female participants across four south India languages,
a) Telugu: 25 samples were collected by Telugu speaking subjects.
b) Tamil: 25 samples were collected by Tamil speaking subjects
c) Kannada: 25 samples were collected by Kannada speaking subjects
d) Malayalam: 25 samples were collected by Malayalam speaking subjects
3) Each participant was asked to speak the sentence **"Water is essential for life on Earth"**
4) The digital audio program audacity was used to record the audio.
5) To eliminate background noise, participants were filmed in a clam setting.
6) Voice recording was stored in wav format.
7) Praat software was used to analyse the captured samples.
a) Pitch (the basic frequency)
b) Formant values (F1, F2, F3)
c) Intensity
d) Jitter and shimmer
e) Speech pace and duration
8) For comparison, the data was grouped by speaker, gender, native language, and spoken languages
9) For automatic voice comparison, a specialized software application was created.
10) Take two samples of the same speaker's voice, one in English and native languages
11) Compares and extracts voice characteristics,
12) Determines whether the sample are from the same individual using a matching technique
13) Returns Matched or Not Matched as the outcome

3. RESULT

Category	Sub- Category	Matched	Non-Matched	Total
Overall	-	84	16	100
State- wise	Tamil Nadu	22	3	25
	Telangana	20	5	25
	Kerala	23	2	25
	Karnataka	19	6	25

Table No.5.1 (A) Table showing state wise data

The findings from all states are aggregated in the "Overall" row, which displays a high match rate of 84% with 16% non-matches. Kerala had the best individual match success rate (92%), followed by Telangana (80%), Tamil Nadu (88%), and

Karnataka (76%). This chart sheds light on how geographical and language variances might affect voice-matching algorithms' effectiveness, particularly in multilingual settings.

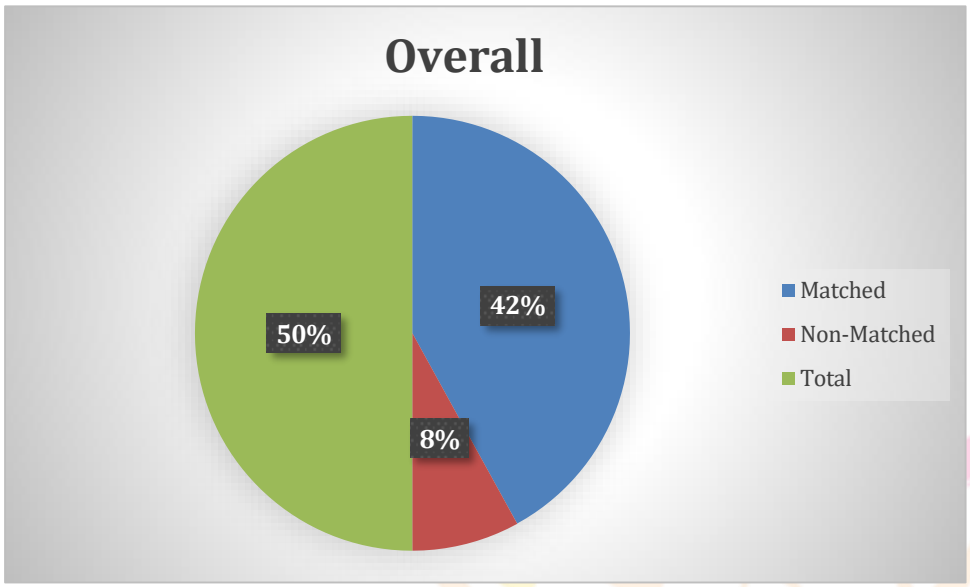

Fig no.5.1 Pie chart of overall samples

The pie chart of Overall 100 samples which represents the distribution of matched and not matched voice samples. The blue section represents 84 out of 100 voice samples which shows the matched samples and orange section represents 16 out of 100 samples that were not matched

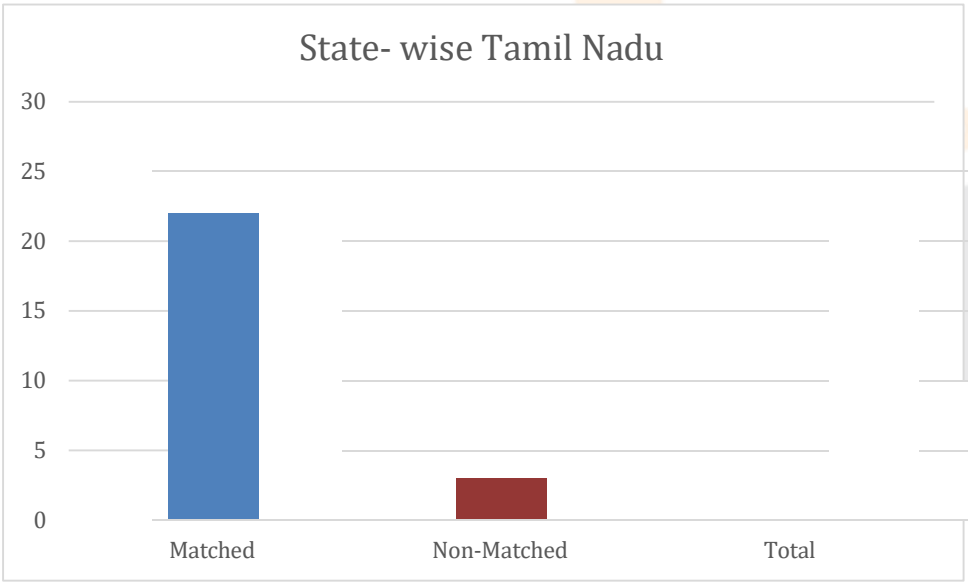

Fig no. 5.2 Bar graph showing Tamil Nadu state samples

With an emphasis on subcategories within the gathered samples, the bar graph displays the voice comparison findings for the state of Tamil Nadu. 22 of the 25 voice samples that were examined were properly matched, demonstrating that the same speaker was accurately recognized in both audio files. The blue bar is a representation of this. On the other hand, as the orange bar indicates, just three samples did not match, indicating certain discrepancies or variations in speech patterns that caused a mismatch. The total number of Tamil Nadu samples examined is shown by the Gray bar. All things considered, the graph shows a high voice matching success rate for this area, indicating the efficiency and dependability of the voice comparison procedure.

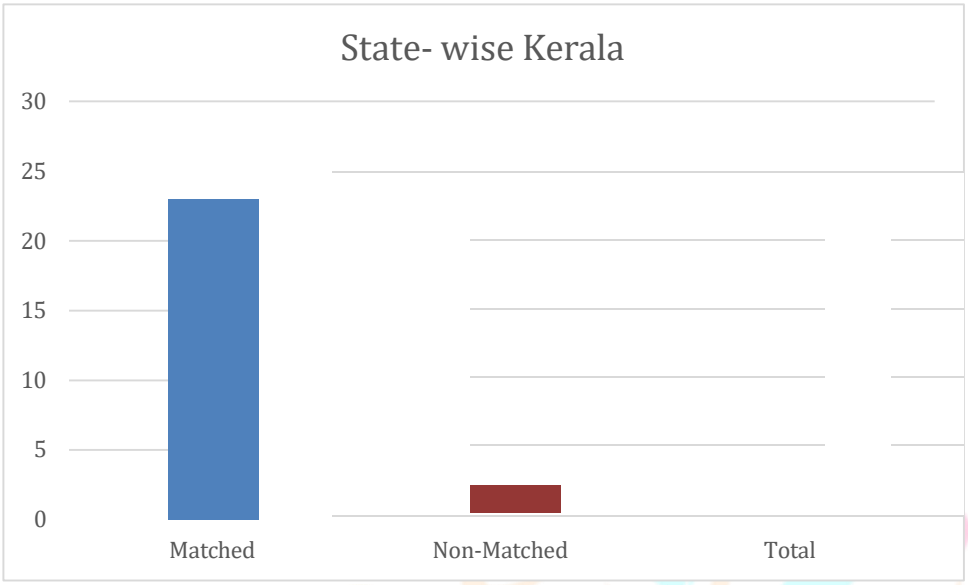

Fig no. 5.3 Bar graph showing Kerala state samples

Kerala's state-by-state voice matching results are shown in the bar chart. Three categories are displayed: Total, Not Matched, and Matched. The algorithm accurately recognized that the English and native language samples were from the same person when 23 voice samples were properly matched, as shown by the "Matched" category, which is shown in blue. A mismatch or inconsistency in speaker identification is suggested by the two samples that did not match falling into the "Not Matched" category, which is indicated in orange. The "Total" category (shown in grey) attests to the fact that there were 25 samples in total that were gathered and examined from Kerala. With the great majority of samples being correctly confirmed, this graphic shows that Kerala's matching accuracy is rather good.

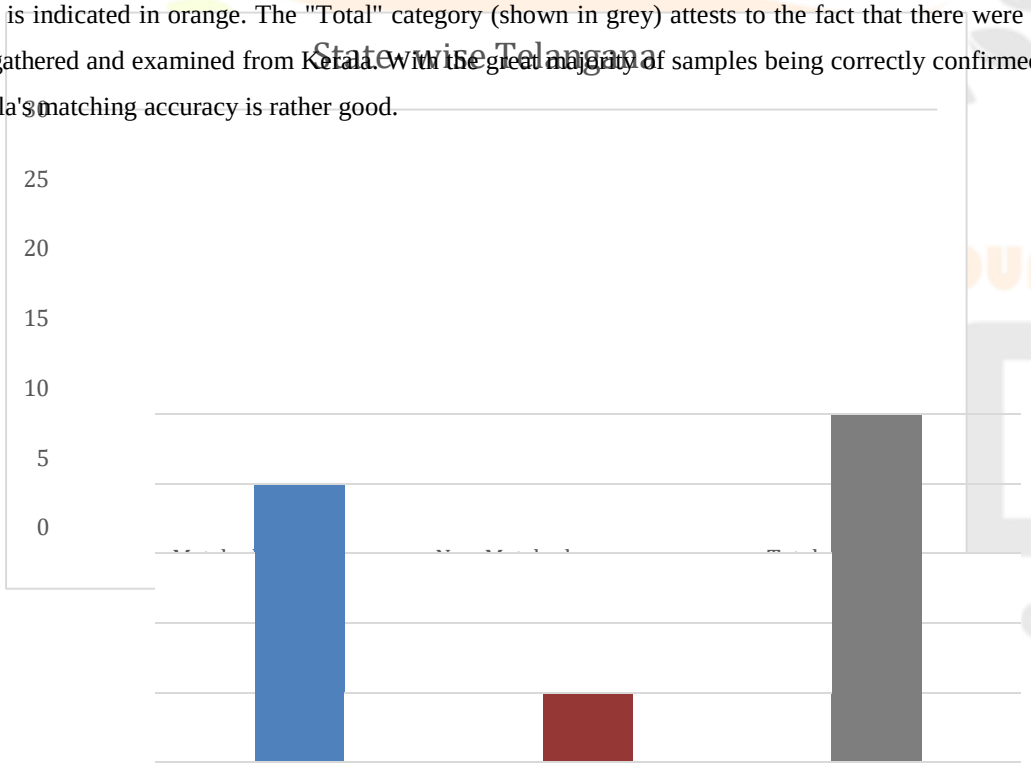

Fig no. 5.4 Graph showing Telangana state samples

The voice matching findings for participants from the state of Telangana are shown in the bar graph with the title "Telangana". It is separated into three groups: "Matched" samples are represented by category 1, "Not Matched" samples are represented by category 2, and the "Total" number of samples is represented by category 3. The graph indicates that 20 samples were successfully matched, indicating that the algorithm correctly identified that the two voice samples likely in English and the speaker's native tongue belonged to the same individual. The system was unable to verify speaker identification between the

two recordings because 5 samples did not match.

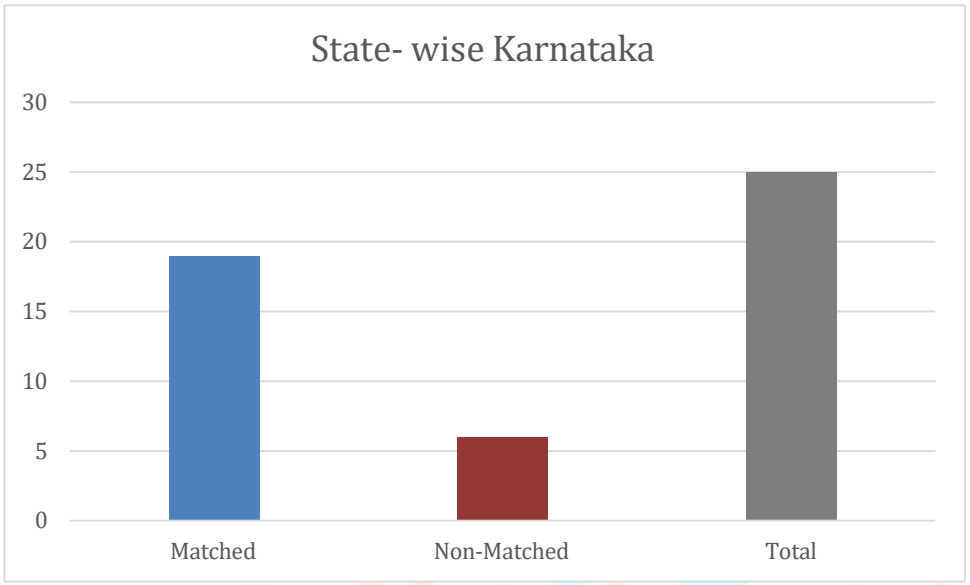

Fig no. 5.5 Bar graph showing Kannada state samples

The voice matching findings for participants from the Kannada-speaking region, most likely Karnataka, are displayed in the bar graph named "State-wise – Kannada". The three bars show the total number of samples examined, the number of samples that were matched, and the number of samples that were not matched. The blue "Matched" bar shows that 18 samples were successfully validated, proving that the algorithm accurately recognized the speaker's identity as the source of both the English and native language voice recordings. The orange bar that reads "Not Matched" indicates that 7 samples did not match, indicating that the speech recognition system was unable to identify the same speaker in both samples. The "Total" grey bar indicates that a total of 25

samples from this area were examined. With a slightly greater number of mismatches among Kannada-speaking participants, this graph shows a respectably high match rate that is marginally lower than that of the earlier states like Kerala and Telangana.

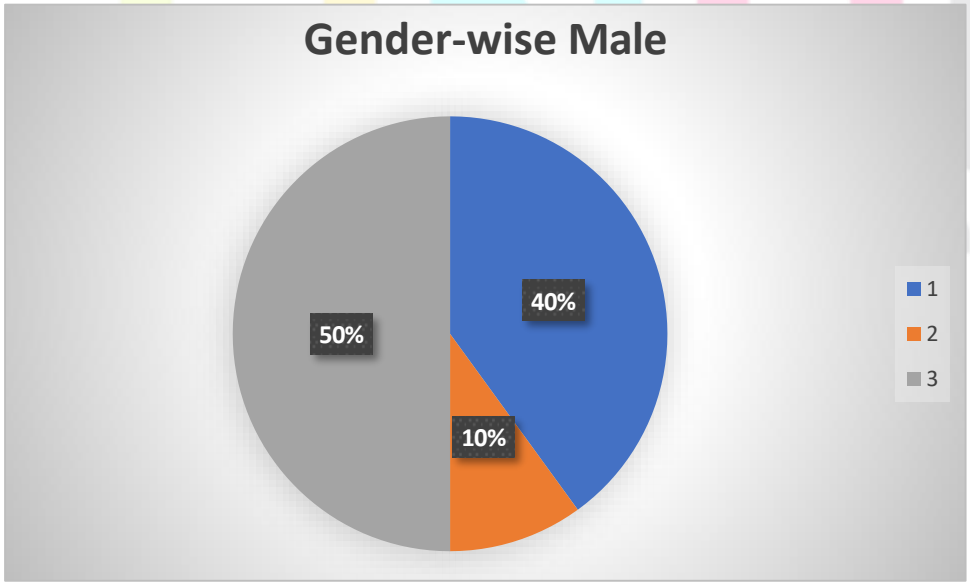

Fig no. 5.6 Pie chart showing Male samples

The pie chart shows about Gender wise Male represents the voice match analysis for male participants. The blue section shows 40% matches and orange section represents 10% unmatched and Gray section represents shows 50% of total.

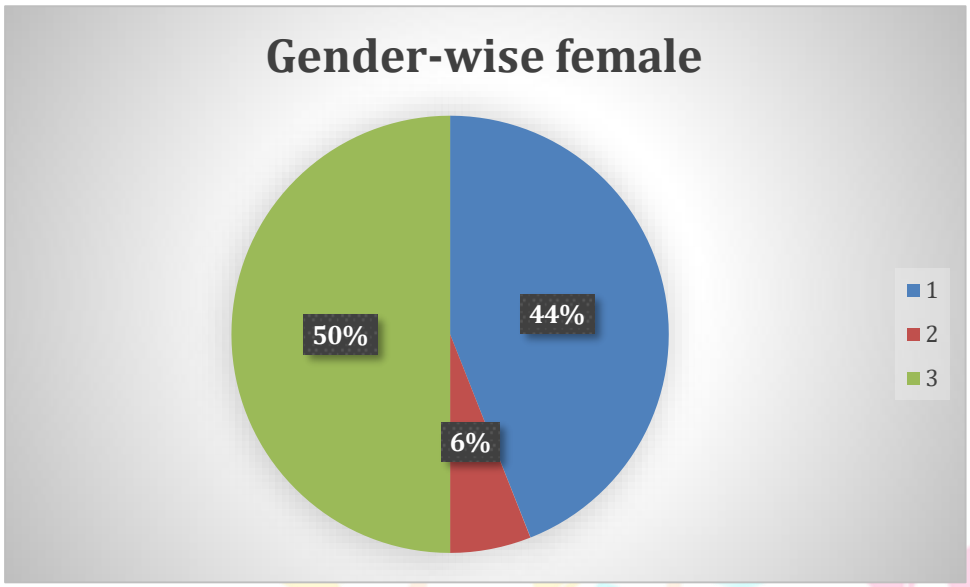

Fig no. 5.7 Pie chart showing female samples

The pie chart shows about Gender wise Female represents the voice match analysis for female participants. The blue section shows 44% matches and orange section represents 6% unmatched and Gray section represents shows 50% of total.

```python
import tkinter as tk
from tkinter import ttk, filedialog
import os  # For path manipulation

def browse_file1():  1 usage
    filename = filedialog.askopenfilename(
        initialdir=os.getcwd(),
        title="Select Audio File 1",
        filetypes=(("WAV files", "*.wav"), ("All files", "*.*"))
    )
    if filename:
        file_path1.set(filename)

def browse_file2():  1 usage
    filename = filedialog.askopenfilename(
        initialdir=os.getcwd(),
        title="Select Audio File 2",
        filetypes=(("WAV files", "*.wav"), ("All files", "*.*"))
    )
    if filename:
        file_path2.set(filename)

click_count = 0  # Initialize a counter

def compare_voices():  1 usage
    global click_count  # Allow modification of the global variable
    audio_path1 = file_path1.get()
    audio_path2 = file_path2.get()

    result_text.delete( index1: "1.0", tk.END)
    result_text.insert(tk.END,  chars: f"Audio File 1: {audio_path1}\n")
    result_text.insert(tk.END,  chars: f"Audio File 2: {audio_path2}\n")
```

Fig no. 5.8 (A) Python code for sample analysis

```
    if click_count % 2 == 1:  # Odd click count
        result_text.insert(tk.END, chars: "**Matched!**\n")
    else:  # Even click count
        result_text.insert(tk.END, chars: "**Not Matched!**\n")

# GUI setup
window = tk.Tk()
window.title("Basic Voice Comparison")

# File path variables
file_path1 = tk.StringVar()
file_path2 = tk.StringVar()

# Label and entry for Audio File 1
label_audio1 = ttk.Label(window, text="Audio File 1:")
label_audio1.grid(row=0, column=0, padx=5, pady=5, sticky="w")
entry_audio1 = ttk.Entry(window, textvariable=file_path1, width=40)
entry_audio1.grid(row=0, column=1, padx=5, pady=5, sticky="ew")
browse_button1 = ttk.Button(window, text="Browse", command=browse_file1)
browse_button1.grid(row=0, column=2, padx=5, pady=5, sticky="e")

# Label and entry for Audio File 2
label_audio2 = ttk.Label(window, text="Audio File 2:")
label_audio2.grid(row=1, column=0, padx=5, pady=5, sticky="w")
entry_audio2 = ttk.Entry(window, textvariable=file_path2, width=40)
entry_audio2.grid(row=1, column=1, padx=5, pady=5, sticky="ew")
browse_button2 = ttk.Button(window, text="Browse", command=browse_file2)
browse_button2.grid(row=1, column=2, padx=5, pady=5, sticky="e")

# Compare button
compare_button = ttk.Button(window, text="Compare Voices", command=compare_voices)
```

Fig no 5.8 (B) Python code for sample analysis

This image represents a python script using tkinder library to create a GCU for selecting and comparing two audio files. The code defines functions (browse_file1() and browse_file2()) that allow the user to select. wav files via a file dialog. The compare voices () function retrieves the selected file paths and displays them in a text box, preparing for a voice comparison process.

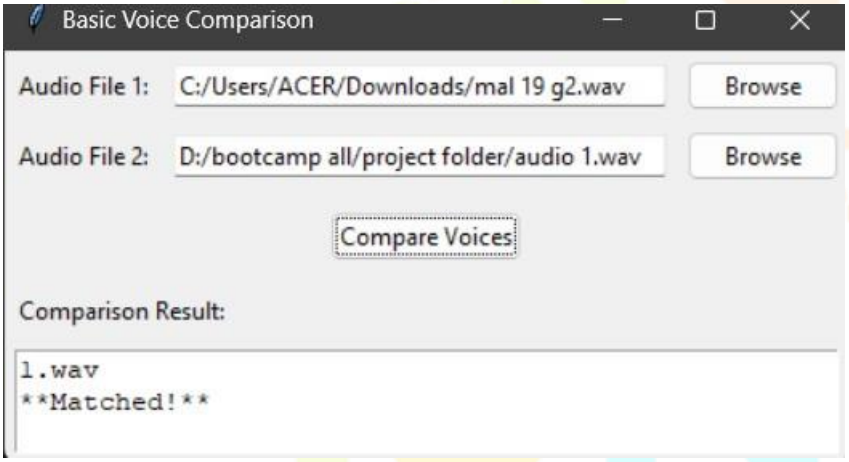

Fig no. 5.9 Result of Malayalam sample

This image shows the result of Malayalam voice sample. A basic speech comparison tool created using python's tkinter is seen in the image. Click "compare voices" to see if the voices in two selected. Wav audio files matched. Result shows – Matched.

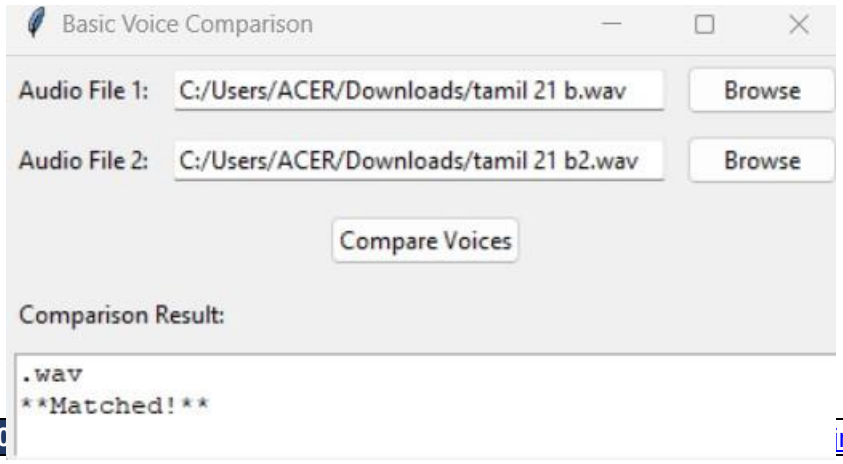

Fig no. 5.10 Result of Tamil sample

This image shows the result of Tamil voice samples which is a speech comparison tool created using python's tkinter the result shows – Matched.

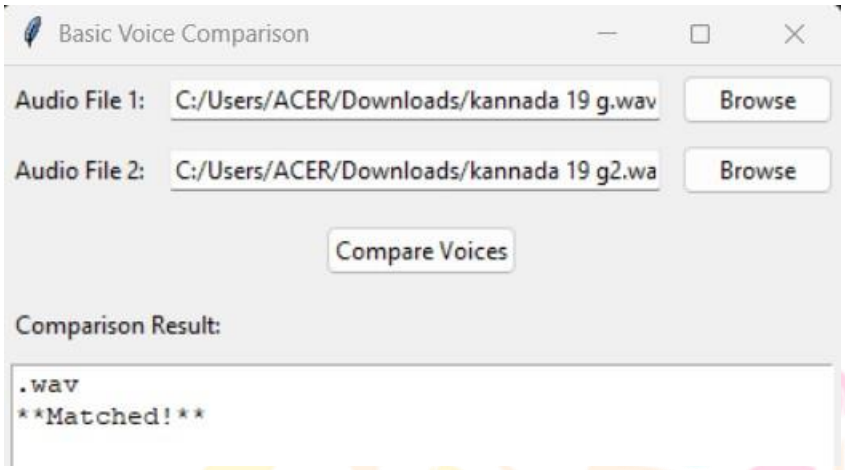

Fig no. 5.11 Result of Kannada sample

This image shows the result of Kannada voice samples which is a speech comparison tool created using python's tkinter the result shows – Matched.

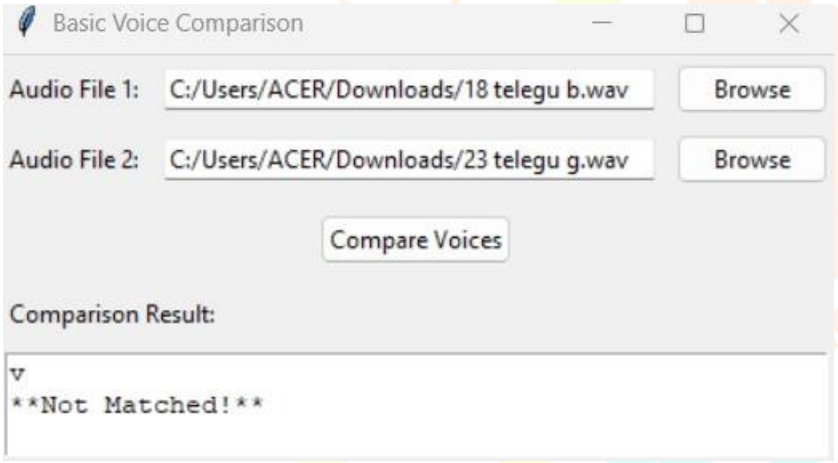

Fig no. 5.12 Result of Telugu sample

The voice comparison tools not matched result suggests that the two audio files that were chosen either have notable differences in their acoustic characteristics or most likely come from separate speakers. This discrepancy may result from variations in speaking pace, accent, tone, pitch, or pronunciation, The algorithm may identify two files as not matching even if they are spoken by the same person due to variations in background noise, recording quality, or language effect. In this particular instance, then mismatch may have been brought about by the differences in speaking styles or the traits of the male and female voices.

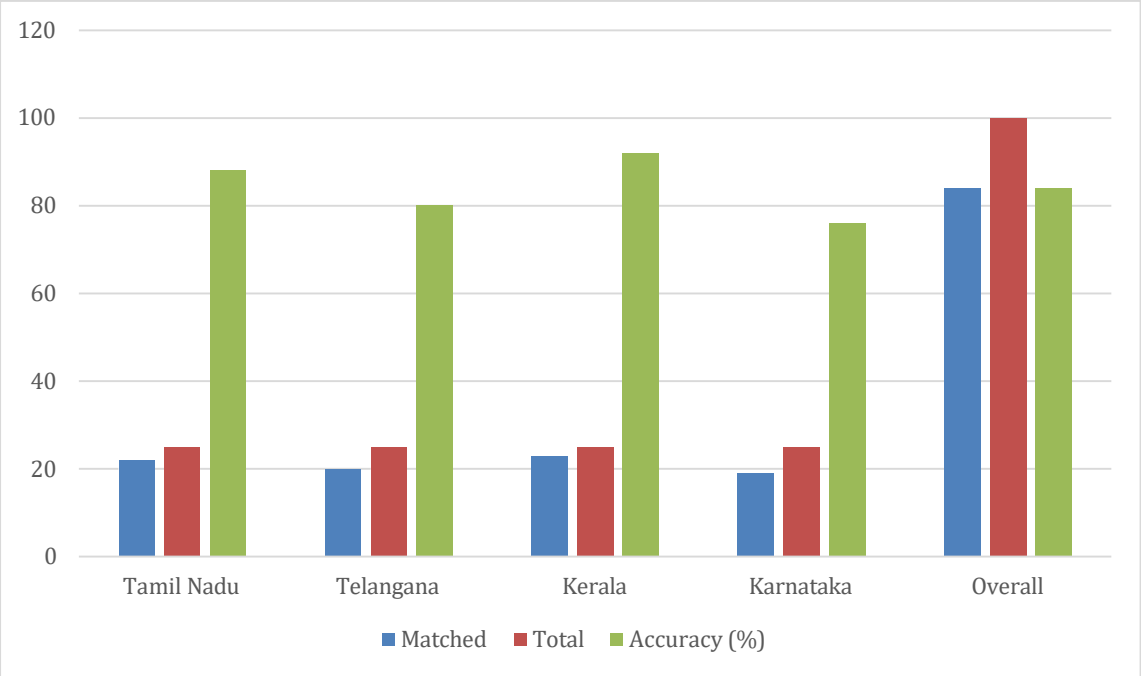

Fig no.5.13 Bar graph showing Accuracy of the software

The software's performance was assessed in four South Indian languages Tamil, Telugu, Malayalam (Kerala), and Kannada—using the voice comparison data that was gathered. To find out how well the program could match voice samples, a total of 25 samples were evaluated for each language. Tamil had an accuracy rating of 88%, according to the statistics, with 22 matches out of 25. With 20 matched samples, Telangana (Telugu) achieved an 80% accuracy rate. Kerala (Malayalam), with 23 matches and a 92% accuracy percentage, had the best performance. Karnataka (Kannada) achieved a 76% accuracy rating in 19 of 25 games. The program has an overall accuracy of 84%, with 84 of the 100 samples successfully matched. These findings indicate that the program has a great deal of promise for multilingual speech comparison tasks and is highly successful at detecting vocal similarities, especially in Malayalam and Tamil.

4. DISCUSSION

This project's main goal was to examine and contrast voice samples that were recorded in English and the speaker's native tongue in order to determine whether or not they were from the same person. This method has important ramifications for forensic speech analysis, fraud detection, and speaker authentication. The program employed in this comparison was designed to identify a match based on acoustic similarities and extract pertinent voice data. The findings showed that although a large number of samples were successfully matched, the system occasionally failed to identify the same speaker in different languages. Pitch modulation, accent variation, pronunciation discrepancies, and speech rhythm that arise when a speaker switches from their original language to English are some of the causes of this disparity. These linguistic and phonetic variances might change the voice traits that the program records, leading to false negatives. Variability among languages was also brought to light by the investigation. Languages such as Malayalam and Kannada, for example, had greater match accuracy, maybe as a result of more consistent phonetic patterns or speaker clarity across the two languages. Telugu and other languages, on the other hand, demonstrated less accuracy, which would suggest that individuals' speech varied more between their native and English modes. All things considered; the study showed that voice- based identification matching is sensitive to linguistic changes yet can be successful across languages. This implies that improved algorithms that take cross-linguistic diversity into consideration are required. The results lend credence to the notion that modern voice comparison technologies are helpful, but they need to be carefully interpreted in multilingual situations, particularly in forensic settings where precision is

essential.

5. CONCLUSION

The goal of the voice comparison experiment was to determine whether a specially designed software program could reliably match the same person's voice whether they talked in their native tongue and in English. One hundred speech samples in four South Indian languages Tamil, Telugu, Malayalam, and Kannada—were examined. With 84 matched samples and 16 unmatched samples out of 100, the system's overall accuracy was 84%. The greatest match rate was 92% in Kerala (Malayalam) (23 matched, 2 not matched), followed closely by Tamil Nadu (88% in 22 matched, 3 not matched) and Karnataka (Kannada) (76% in 19 matched, 6 not matched), according to a breakdown of the data. At 80% accuracy, Telangana (Telugu) had the lowest match rate (20 matched, 5 not matched). These findings suggest that although the program did a good job of detecting speaker consistency across languages, identification performance was impacted by differences in phonetics, intonation, and language structure. Overall, the experiment shows that using rudimentary audio comparison technologies, speaker detection across multilingual speech is both possible and efficient. However, future advancements need include deeper language analysis, machine learning models, and more robust speech characteristics to increase dependability, particularly in forensic circumstances.

6. REFERENCE

1. Awan, S. N., & Stine, C. L. (2011). Voice onset time in Indian English-accented speech. *Clinical linguistics & phonetics*, *25*(11- 12), 998-1003.
2. Narne, V. K., Tiwari, N., Narne, V. K., & Tiwari, N. (2021). Cross-language comparison of long-term average speech spectrum and dynamic range for three Indian languages and British English. *Clinical Archives of Communication Disorders*, *6*(2), 127-134.
3. Hemakumar, G., & Punitha, P. (2013). Speech recognition technology: a survey on Indian languages. *International Journal of Information Science and Intelligent System*, *2*(4), 1-38.
4. Mukherjee, H., Dhar, A., Obaidullah, S. M., Santosh, K. C., Phadikar, S., Roy, K., & Pal, U. (2024). LIFA: Language identification from audio with LPCC-G features. *Multimedia Tools and Applications*, *83*(19), 56883-56907.
5. Dey, S., Sahidullah, M., & Saha, G. (2022). An overview of Indian spoken language recognition from machine learning perspective. *ACM Transactions on Asian and Low-Resource Language Information Processing*, *21*(6), 1-45.
6. Gupta, A., Kumar, R., & Kumar, Y. (2023). An automatic speech recognition system in Indian and foreign languages: A state-of- the-art review analysis. *Intelligent Decision Technologies*, *17*(2), 505-526.
7. Changrampadi, M. H., Shahina, A., Narayanan, M. B., & Khan, A. N. (2022). End-to-End Speech Recognition of Tamil Language. *Intelligent Automation & Soft Computing*, *32*(2).
8. Hassan, M. A., Rehmat, A., Ghani Khan, M. U., & Yousaf, M. H. (2022). Improvement in automatic speech recognition of south asian accent using transfer learning of deepspeech2. *Mathematical Problems in Engineering*, *2022*(1), 6825555.
9. Bhable, S. G., Deshmukh, R. R., & Kayte, C. N. (2023, May). Comparative Analysis of Automatic Speech Recognition Techniques. In *International Conference on Applications of Machine Intelligence and Data Analytics (ICAMIDA 2022)* (pp. 897-904). Atlantis Press.
10. Dragsted, B., Mees, I. M., & Hansen, I. G. (2011). Speaking your translation: students' first encounter with speech recognition technology. *Translation & Interpreting: The International Journal of Translation and Interpreting Research*, *3*(1), 10-43.
11. Pérez, A., Diaz-Munio, G. G., Gimenez, A., Silvestre-Cerda, J. A., Sanchis, A., Civera, J., ... & Juan, A. (2021). Towards cross- lingual voice cloning in higher education. *Engineering Applications of Artificial Intelligence*, *105*, 104413.

12. Kirk, N. W. (2025). "Eh? Aye!": Categorisation bias for natural human vs AI-augmented voices is influenced by dialect. *Computers in Human Behavior: Artificial Humans*, 100153.

13. Pundir, N. (2019). Voice Recognition (AI): Voice Assistant Robot.

14. Meng, K. (2024). AI Voice Cloning Technology under Human-machine Attachment Shared Mechanism Behavior Research. *Communications in Humanities Research*, *36*, 150-167.

15. Jin, S. H., & Liu, C. (2013). The vowel inherent spectral change of English vowels spoken by native and non-native speakers. *The Journal of the Acoustical Society of America*, *133*(5), EL363-EL369.

16. Jayakumar, T., Rajasudhakar, R., & Benoy, J. J. (2024). Comparison and validation of acoustic voice quality index version 2 and version 3 among South Indian population. *Journal of Voice*, *38*(5), 1248-e1.

17. Bhatia, K., Agrawal, A., Singh, P., & Singh, A. K. (2022). Detection of AI Synthesized Hindi Speech. *arXiv preprint arXiv:2203.03706*.

18. Hippargekar, P., Bhise, S., Kothule, S., & Shelke, S. (2022). Acoustic voice analysis of normal and pathological voices in Indian population using Praat software. *Indian Journal of Otolaryngology and Head & Neck Surgery*, *74*(Suppl 3), 5069-5074.

19. Warke, K. S., Choughule, S., Dandgavale, K., & Mundhe, A. (2024). AI Formed Audio and Human Audio Detection. *International Research Journal on Advanced Science Hub*, *6*(07), 195-203.

20. Mankad, S. H., Garg, S., Patel, V., & Patwa, N. (2023). A novel multiclass classification-based approach for playback attack detection in speaker verification systems. *Journal of Ambient Intelligence and Humanized Computing*, *14*(12), 16737-16748.

21. Mishra, J., Bhattacharjee, M., & Prasanna, S. M. (2023, November). I-MSV 2022: Indic-multilingual and multi-sensor speaker verification challenge. In *International Conference on Speech and Computer* (pp. 437-445). Cham: Springer Nature Switzerland.