Voice Assistant (Emotion Detection) Using Python

Manasvi S
Department of Computer Science and Engineering
Alliance University
Bengaluru, India
smanasvibtech22@ced.alliance.edu.in

DoreSwamy S B
Department of Computer Science and Engineering
Alliance University
Bengaluru, India
sdoreswmybtech22@ced.alliance.edu.in

Nitya
Department of Computer Science and Engineering
Alliance University
Bengaluru, India
nnityabtech22@ced.alliance.edu.in

Poornima
Department of Computer Science and Engineering
Alliance University
Bengaluru, India
smanasvibtech22@ced.alliance.edu.in

Darshan R
Department of Computer Science and Engineering
Alliance University
Bengaluru, India
rdarshanbtech22@ced.alliance.edu.in

DR Sridhar Devarajan
Department of Computer Science and Engineering
Alliance University
Bengaluru, India
sridhar.devarajan@alliance.edu.in

ABSTRACT

This research on the Emotion-Detecting Voice As- sistant Project details the advancements , additional testing and refinement undertaken to enhance the emotions recognition capabilities. Initial design covered in the first report , this phase focuses in implementing emotion classification model , it detects the range of emotion detection and improving the accuracy and responsiveness of voice assistant.By using the combination of advanced machine learning techniques and real- time processing optimizations , we developed the model to handel diverse audio inputs . There are some key points which has to be improved which includes adjustments in audio feature extraction , optimization of classification algorithms and an expansion of the training dataset to account for subtle variations in emotion expression . As a result the models accuracy has been improved. In addition to technical improvements , this report examines user feedback and the ethical considerations linked with emotion detection in AI. This report on emotion-detection provides depth analysis of the enhancements made to improve the assistants accuracy, overall usability .After completing the first report we also implemented this model by technical and new approachments . These advancements position the emotion- detecting voice assistant as a promising tool in all the services , where understanding the emotions are playing main role . Future work will focus on real-time development , learning capabilities . The implementation provides advanced feature extraction techniques through the library , capturing detailed audio features such as pitch,tone and emotions, which are essential for classifying emotions Introduction

INTRODUCTION

Recent advances in the spheres of both machine learning and artificial intelligence have defined us with new relations toward technologies. Literally to put and carry with oneself, voice assistants were created as devices that can be useful in life; however, assisting in information searching, they also represent devices for interactive experiences. An important feature in real human communication is that most voice assistants are incapable of detecting-and, hence, responding to the emotional intent of the user comments. The limitation can lead to misinterpretations or responses that may seem unaffected and lacking, but especially in situations when such sensitivity is critical, such as while delivering customer service or for support to mental health. This project tells of the creation of a voice assistant that will recognize and react to user emotion through recognizing a need for even more intuitive interactions. It is made by Python, the assistant working towards completely closing the gap between emotional intelligence and artificial intelligence by providing more relevant and human-centered interactions. The assistant may be very engaging to the user experience by adopting its responses to the user's mood through emotional clues identified within speech while responding with speech, so that it can make the user feel special or feel comfortable. The voice assistant acquires popularity due to its hands-free provision about the access to it, the management of tasks, and well-manered services. These assistants do almost everything from A to Z. Like playing music and displaying reminders to quote quickly to be able to respond to inquiries, detecting emotions have absolutely transformed the way mankind interacts with technology. Because of easy use through voice-activated devices, there is an overwhelming demand for easy, accessible, and flexible voice assistant solutions and most especially that can be specialized towards specific needs and applied environments. Such assistants made daily life very easy and this, at times was taking most of the role unconsciously. In the past few years, there have been lots of research studies focused on speech emotion recognition as an intent to classify emotions, primarily working in two areas: some try to seek more representative features or fusion of different features for emotion recognition in speech, and others concentrate on the kind of machine learning models used, such as GMM, SVM, and even deeper learning methods like CNN, Long short-term memory (LSTM), in order to enhance the performance in emotion recognition from speech. There exists an actual term, used in the field of artificial intelligence, for emotion recognition from speech inputs-called Speech Emotion Recognition (SER). The manner in which we speak essentially communicates to our psyche; states of emotions like joy or sadness, anger, surprise, or fear may all be conveyed in terms of changes in tone, pitch, loud-ness, pace, and rhythm. Speech Emotion Detection is thus developing a model that detects emotions based on voice assistant based on proper detection of the user's emotional state through analysis of these vocal characteristics. Unique characteristics corresponding to distinct emotions form the basis for Speech Emotion Recognition technology. For instance, neutral could be exemplified by a softer, slower tone, but higher pitch and louder volume often represent or excitement. These features can then be tilded to the machine learning algorithms. This includes ease of use for voice-activated devices, which forms a significant demand for easily accessible and adaptable voice assistant solutions, especially those that can be tailored to fit specific requirements and settings. It has earned its

Autonomy in making daily life very easy because it knows it is taking up most of the role, unknowingly. In the last few years, a significant amount of research has focused on speech emotion recognition to classify emotion. Based on these works, most studies are mainly concentrated in two aspects: more representative features or fusion of different features for emotion recognition in speech ; machine learning models, including common emotion classifiers like Gaussian Mixture Model (GMM) and Support Vector Machine (SVM), or deep learning methods like Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM), etc., to improve performances of emotion recognition from speech.The term specifically existing within the domain of artificial intelligence is known as Speech Emotion Recognition or SER, which concerns the task of emotion recognition from speech inputs. The way one speaks is very much n a position to influence its emotive

B. SYSTEM ARCHITECTURE

High-LevelComponents:

1. Interface Front end

- Voice Input Interface: User sound recording for processing, in cellphone application, laptop, or smart device.
- Response Output Interface: Using TTS in order to offer the answer, supposedly modulate the tone to reflect emotion.

2.Works at the Ground Emotional Introduction Services

- Performs audio featuresextraction, including vocal features like MFCCs and vowels,and feels classifications using previously trained styles.
- There is use of a neural network or SVM with trained categorized sensory facts for example, RAVDESS or TESS.
- External Integration Office: It deals with all the connections to the outside structures including calendar or weather APIs to execute specific commands or represent statistics.
- Harvesting and Analysis: it is based on logs of person interactions, sensory attributes and system settings for evaluation (with person consent) to make the machine more efficient and scalable. 3. Data warehouse and processing
- Customer facts warehouse: This warehouse holds customer profiles, historical interactions, and emotive features if customers allow so.
- Storage and Maintenance: It collects and periodically updates the model of emotion recognition styles according to novel information and user feedback 4. Infrastructure factor
- API Gateway: It is in charge of all of the incoming requests from the front, and it can successfully route to the correct initiatives and enforces pricing policies.
- Load Balance: It distributes incoming requests smoothly to backend services to handle accelerated traffics. Business process variant Examples:

1. User Interaction: The user speaks to the assistant. The Voice Input Interface data audio.
2. Sentiment recognition and ASR: The emotional tracking feature analyzes voice functionalities or classifies the user's emotional state. The voice recognition feature transcribes the audio for further processing. emotion. For example, softer and more slower in tone of voice is usually when viewed as a neutral emotion, whereas high pitch and loud volume mostly convey excitement emotion. The problems mentioned and the same, that is removal of these problems, form the Emotion Detection system with a great model trained hardly to identify different varieties of emotions in real time.

II. LITERATURE REVIEW

A.• International Research Journal of Modernization in Engineering Technology and Science Volume:03/Issue:05/May- 2021

The first chatbot was made in 1960 which confirmed the start of improvement in digital assistants Considering current product many new era advanced and maintain advancing. Speech popularity is one of the techniques which allows to enhance the capability of digital assistants. Now there are many special languages available to assist humans and it continues enhancing. For emotion reputation researchers work for many years and had been capable of successfully develop. Different methods and techniques are evolved like Bayesian networks ,Gaussian Mixture models and Hidden Markov models and deep neural networks. Speech popularity and emotion popularity each were advancing in their very own field. Many digital assistants like Alexa, Siri are used in many places. Their cutting- edge need is to put into effect emotion reputation approach with a purpose to help to recognize our emotions. The development is referred to as Emotion AI. Prototypes and business products for emotion driven assistants already exist as an example Beyond Verbal's voice recognition app, Emotion AI for speech by using Affective and the linked domestic VPA Hubble.

B.• International journal cyber behavior Volume 13 • Issue

Emotion detection through voice and speech reputation has been drastically studied due to its programs in fields like healthcare, customer support, and interactive AI. Literature on this subject matter explores how feelings are conveyed thru acoustic capabilities such as pitch, tone, and rhythm. Vari- ous techniques, together with device mastering (e.g., Support Vector Machines, Convolutional Neural Networks) and deep getting to know fashions, had been applied to seize these complicated patterns. However, demanding situations persist, specifically with variability in emotional expression through- out distinctive languages and cultural contexts. This has led to analyze specializing in constructing more generalized and strong systems for accurate emotion reputation.

C • International Journal of Research in Engineering and Sci- ence (IJRES)ISSN (Online): 2320-9364, ISSN (Print): 2320- 9356 www.ijres.org Volume 10 Issue 2 Paper's literature review focuses on recent developments in virtual personal assistants (VPAs) and their underlying technologies, including artificial intelligence (AI) and natural language processing (NLP).) including the so. Leading com- mercial examples such as Siri, Amazon Alexa, and Google Assistant, highlight how industry leaders are using VPAs to facilitate human-machine interactions The review introduces the development of a Python-based desktop voice assistant that works on Windows systems. Speech Signal-Based Modelling of Basic Emotions to Analyse Compound Emotion: Anxiety.

D • International Research Journal of Engineering and Tech- nology (IRJET) Volume: 09 Issue: 05 — May 2022 The literature review in the paper ”Voice Assistant Using Python and AI” includes studies on speech recognition, natural language processing (NLP), and machine learning techniques needed to create virtual assistants: Speech recognition: Key research by Atal Rabiner et al examines speech signal pro- cessing to convert analog signals to digital waveforms, which are important for accurate speech recognition often Mel- Frequency Cepstral Coefficients (MFCC) and Hidden Markov for analyzing and classifying command utterance Models (HMM) and other techniques. NLP and Machine Learning:

NLP enables virtual assistants to process and interpret human speech, facilitating commands and responses in natural language. Key techniques include Dynamic Time Warping (DTW) and Neural Networks, which support functions such as keyword extraction from sentiment analysis. Functionality and accessibility: The study notes the functionality of accessibility, such as assisting the elderly or visually impaired with voice guidance in daily tasks .

e.g reading the news or monitoring home appliances note. Voice assistants like Amazon Alexa are examples of these capabilities.

III. SYSTEM DESIGN

A Problem Definition

Develop an intelligent speech emotion recognition system that can accurately recognize and segment human emotions from spoken language, enabling devices to empathically understand and respond to users’ emotional states Have a voice assistant who can recognize and respond to a person’s emotions especially based on verbal feedback. Traditional voice assistants respond in an unbiased tone and do not account for the emotional state of the user, may be impersonal or have side-by-side interactions This assistant will enhance the user’s interest through emotions such as excitement , sadness, anger, or neutrality

The purpose of this selection is to:

1.Confirm the sensitivity of the user’s language.

2.Respond in a manner that is based on well-understood feelings and on exchange information.

3.Data garage and processing

• User facts garage: It stores consumer profiles, past interactions, and emotional attributes if customers consent.

• Storage and maintenance: Collects and periodically updates emotion recognition fashions based totally on new data and user remarks 4.The infrastructure thing .

• The API Gateway:Manages all requests from the front to lower back, efficiently routes to suitable initiatives, and enforces pricing regulations.

• Load Balance:It calmly distributes incoming requests to backend offerings to deal with accelerated traffics. Business manner version

Examples: 1.User Interaction: The user talks to the assistant. The Voice Input Interface data audio.

2.Sentiment recognition and ASR: The emotion tracking characteristic analyzes voice functions and classifies the emotional state of the user. The voice recognition characteristic converts the audio into tran- scripts for further processing

.

B. SYSTEM ARCHITECTURE : High-Level Components:

Frontend Interface

• Voice Input Interface: Captures in- dividual audio for processing, both on a cellular app, laptop, or clever device.

• Response Output Interface: Uses Text-to- Speech (TTS) to supply responses, doubtlessly modulating tone to reflect empathy.

2.Works at the Ground Emotional Introduction Services

• Processes audio, extracts vocal features (e.G., MFCCs, vowels), and classifies feelings the use of for- merly educated fashions.

• A neural network or SVM skilled on categorized sensory facts (e.G., RAVDESS or TESS) can be used.

• External Integration Office: Manages connections to outside structures which include calendar or weather APIs to execute unique commands or provide statistics.

• Harvesting and AnalysisLogs person interactions, sensory attributes, and system settings for evaluation (with person consent) to make the machine more efficient and scalable.

3.Data garage and processing

• User facts garage: It stores consumer profiles, past interactions, and emotional attributes if customers consent.

• Storage and maintenance: Collects and periodically updates emotion recognition fashions based totally on new data and user remarks 4.The infrastructure thing

• The API Gateway:Manages all requests from the front to lower back, efficiently routes to suitable initiatives, and enforces pricing regulations.

• Load Balance:It calmly distributes incoming requests to backend offerings to deal with accelerated traffics.Business manner version

Examples: 1.User Interaction: The user talks to the assistant. The Voice Input Interface data audio.

2.Sentiment recognition and ASR: The emotion tracking characteristic analyzes voice functions and classifies the emotional state of the user. The voice recognition characteristic converts the audio into tran- scripts for further processing.

System Impelementation:

It's interesting how it puts audio analysis and natural language together.gauge processing, and interactive feedback. I built this python assistant that listens to voices, recognizes

have emotional tones, and responds accordingly. Her ewe take

overview of the architecture and in depth functionality of thissystem.1. Audio compression and feature extraction The first act this voice assistant does is a scan

There are pre-recorded audio files to be uploaded by the user.

They use such a code instead for the real-time voice recording.It uses the librosa library, a Python package specially Designed for the analysis of music and audio, breaking

audio breaks down into parts. The characteristics function as triggers that Translate in the voice of the user their feelings. Some

Among the characteristic features of sound are: RMSLoudness This

It measures the amount of energy in the sound and corresponds it to mood. High energy generally indicates happiness

or rage, a low energy may symbolize peace or melancholy.Tempo; (technique or speed fast tempo - emotional) be: state like excitement or arousal. Slow speed refers to more stress or calm. Voice: Repeat the voice for Know what you are singing. Higher notes can be happy or cheerful and lower notes standing for sad or peace.Spectrum Contrast : This measures the difference between the high and low intensities of frequency spectrum to check if the one that is emphasized is sharp. dimness or lack of light, often related with bright emphasis or. Zero crossing rate: This is a count

of how many times the audio signal becomes frompositive to Lower or vice versa. Increased zero crossing rates tend to accompany louder or more animated voices, while

Lower rates in general also are associated with quieter speech. All

It is this amalgamation which brings an emotional picture It is a thought provoking audio analysis nexus natural language processing, and interactive feedback. I built this

voice assistant from Python which listens and recognizes voices Emotional tones, it responds accordingly. Here we take In the structure and the depth of functionality of this,

System. 1. Audio Compression and Feature Extraction: The first thing this voice assistant does is analyze the user audio inputs. Audio recorded in advance, is not This code is utilized in place of a live voice recorder. This library, librosa, is a Python package specifically Use for music and audio analysis, which breaks audio into components. The features act like cues that

Decode the emotions carried in the voice of the user. Some

The most significant among the critical audio characteristics are: RMS Loudness This Quantifies the amount of energy in a sound and is related it to mood. High energy typically means happiness -or anger while low energy tells of either peace or misery.

Tempo (speed) A fast tempo is an emotionaSuch as excitement or arousal. Low speed refers to More tension or serenity. Voice: Follow the voice toUnderstand what you are singing. Higher notes can be happy or merry, and lower notes sound sad or It grows in peace.Spectrum Contrast: measures difference Between the high and low intensities of the frequency

spectrum, to identify whether it contains a sharp one is emphasized or not brightness, usually accompanied by high

Tonal accent or emphasis. Zero crossing rate: This is a count

Number of countercrossing of audio signal from positive to Negative, or vice versa. High zero crossing rates mostly to accompany stronger or more excited voices, while

Lower rates generally tend to occur with quieter speech. The voice assistant converts spoken words to text at a 80-85For good responses, the assistant maintains a strong ability to respond appropriately that has an empathetic touch to it, with user checks produced accurate consistency in their responses 85The random forest model gives weightage scores to capabilities,

indicating features that were most active in type of sensitivity. The most significant level of perfection, nearly 85-ninetyPercent The emotive popularity component is highly effective in detecting the emotional tone of the person approximately

problems are:MFCCs: In line with expectation, MFCCs took the highest contributing feature in emotion popularity, for capturing the spectral properties of speech

sign Tone: The voice of the speaker was also one of the most important, primarily

for characterizing the difference between positive and bitter emotions. High

pitches are usually associated with happiness, whereas low

pitches tend to be aligned with sad and angry emotions. When the frequencies of predicted emotions were plotted, a prevalence of happiness over anger was seen. This is in line with the RAVDESS database, which has a higher proportion of happy and angry samples than sad and neutral samples. Furthermore, the model did not perform well with neutral samples probably because these have subtle features that could not easily be differentiated by the model.

VI. CONCLUSION AND FUTURE ENHANCEMENTS

The design of a voice assistant using Python has a capability to detect emotions, and it has promising results on the system for just recognition and classification of human emo- tions from spoken language.

The design of an intelligent voice assistant that is capable of recognizing inherent emotions of a user is a major step enhancement in the understanding of the humanity-computer interface.By integrating natural language processing and emotion recognition, he not only perceives what language was uttered but rather comprehends the feeling that was intended. This amalgamation helps in getting out

messages that are rather gentle and more susceptible making the relations even more useful and caring. In terms of the

emotions recognition, the voice assistant can personalize its reply and acoustic which can lead to improved customer satisfaction and higher engagement.

FUTURE ENHANCMENT

1.First, More Emotion Categories for Future Improve-

mends : To better describe what users really feel, future versions can detect more sophisticated emotions such as surprise, fear, surprising, cunning, anxiety, or sarcasm in addition to the more common emotions like happy, sad, and furious. All of these emotions can be used to produce excellent models till now we used it for only three emotions.

2.Contextual Emotion Recognition: Giving the model context-

awareness, the model will then make sense of better emotions by picturizing on previous encounters and context, allowing them to differentiate emotions that might sound similar but have different meanings, there will be cases when pitch doesn't help and it looks like 2 or more emotions so in that time detecting will be easier by contextual awareness.

3.Cross-Cultural and Multilingual Emotion Recognition: Adding the ability to identify emotions in a variety of anguages and many country people will be having.different accents, accents would increase the assistant'accessibility worldwide and allow for more cross-cultural. Emotional baselines and personalization: By gradually learning from each user, the assistant might establish baseline with their typical emotions and tone. The personalize approach would enable it to detect the emotions. These come together to compose a rhetorical photo. Spectrum Contrast: This is obtained by high and low intensities of the frequency spectrum, saying if a sharp one is made prominent or not brightness, which is basically associated with loudness. The subscript for the permeability of vacuum μ_0, and other common scientific, is zero with subscript formatting, not lowercase letter "o". In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical. sentence is punctuated within the parentheses.) A graph within a graph is an "inset", not an "insert". The word alternatively is preferred to the word "alternately"(unless you really mean something that alternates).

ACKNOWLEDGMENT

The satisfaction that comes at the moment of successful completion of the task is left incomplete without mentioning people who made it possible whose continuous guidance and encouragement crowns all efforts with success.

We avail this opportunity to express our deep sense of gratitude and hearty thanks to the Management of Alliance University, for providing world class infrastructure, congenial atmosphere and encouragement.

We take this opportunity to express our gratitude to

Dr.Reeba Korah, Dean, ACED and Dr. A.Ezil Sam Leni, Head of the department of CSE/IT, for their encouragement and cooperation at various levels of the project.

We are very thankful to our Guide Dr. Sridhar Devarjan and Project Coordinators Dr. M. Selvam and Dr. Lopa Mandal, Department of Computer Science and Engineering &

Information Technology for their sustained inspiring guidance and cooperation

throughout the process of this project. Their wise counse and valuable suggestions are invaluable.

We express our deep sense of gratitude and thanks to the teaching and non-teaching staff

At our department who stood with us during the project and helped us to make it a successful venture. We place highest regards to my parents, my friends and well-wishers who helped a lot in making the research of this projectReferences The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2].

REFERENCES

[1] S. Treponemal and S. Natrajan, "Transport Phenomena of Semiconductor Journal of Medical Physics, Vol. 42, No 5, pp. 421-425, May 2005.

[2] P. Banerjee, M. Haldar, D. Zaretsky and R. Anderson, "Overview of 4 Compiler for Synthesizing MATLAB Programs onto FPGAs," IEEE Transactions on Very Large Scale Integration (VLSI) System, Vol. 12, No 3, pp. 312-324, March 2004.

[3] J. Jores (2006), "Contact Mechanics," Cambridge University Press, UK, Chapter 6, pp. 144-164.

[4] C. Rovers Eds., "Recent Advances in DSP Techniques," 2nd ed., Taylors Frances Group, USA 2006.

[5] R. Smith (2008), "Contact of Cylindrical Surfaces". Available Online at: http://www.casphy.cenm.edu/homepage.html J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[6] H. D. Cheng, "Image Features Extraction using Volterra Filters," Proc. of 5th IEEEInternational Conference on Machine Vision and Artificial Intelligence, China, pp42-57, October 2009.

[7] A. K. Barnard, “A Study of Stereo Matching Algorithms for Mobile Robots,” A Thesis Report for University of Bath, U.K., 2005.

[8] J. P. Williamson, “Non-Linear Resonant Granit Devices,” US Patent 3 624 12, July 16, 1990.

[9] Motorola Semiconductor Data Manual, Motorola Semiconductor Products Inc., Phoenix, AZ