**DESIGN,DEVELOPMENT AND ACOUSTIC ANALYSIS OF NORMAL AND LOMBARD SPEECH TAMIL CORPUS**

***RESEARCH INTERSHIP FOR STUDENTS OF OTHER INSITUTIONS***

**DEPARTMENT OF INFORMATION TECHNOLOGY**

**SRI SIVASUBRAMANIYA NADAR COLLEGE OF ENGINEERING**

**KALAVAKKAM, CHENNAI 603 110**

***Submitted by***

**NAVEEN K**

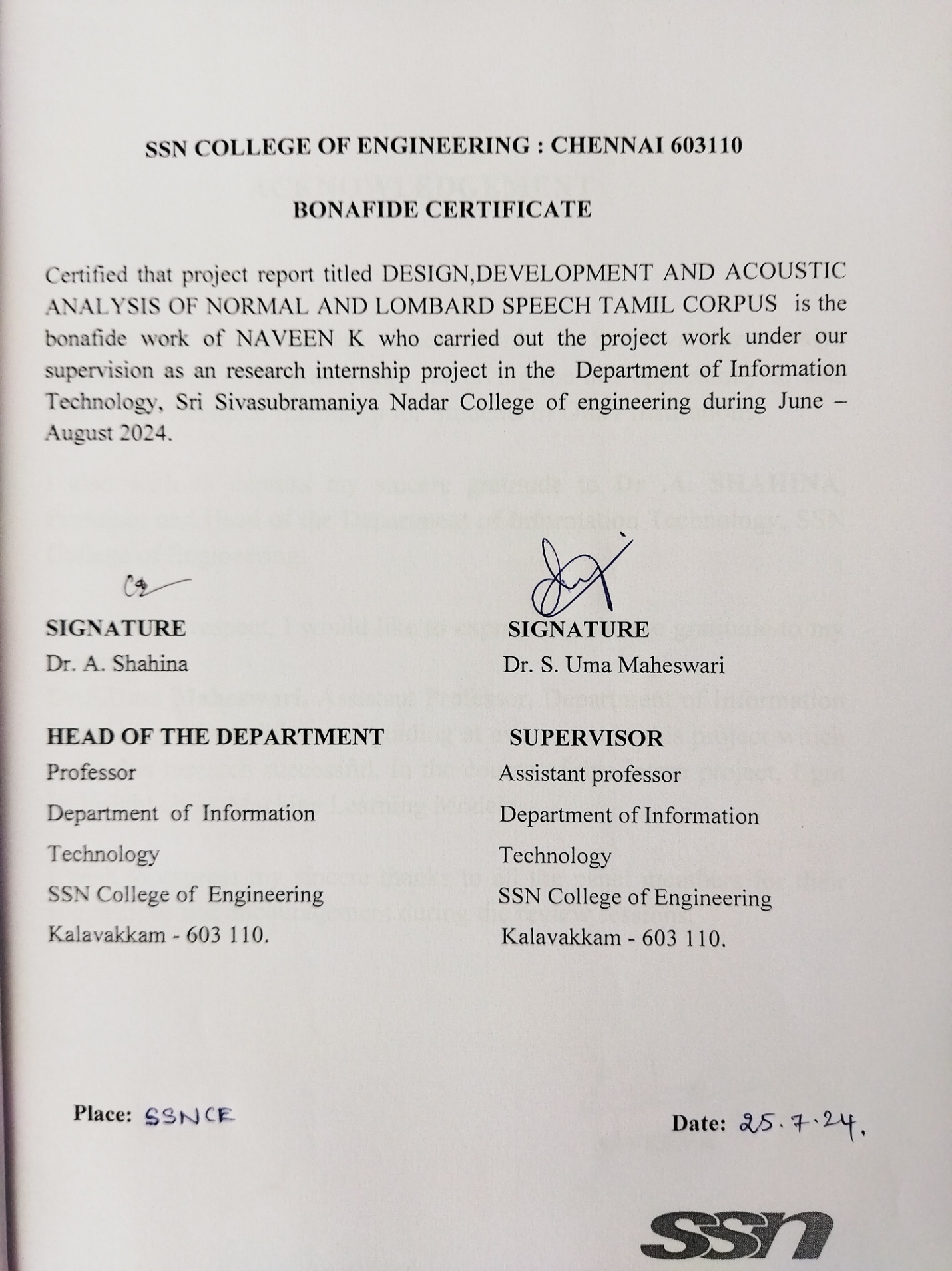
**INFORMATION TECHNOLOGY**

**ANJALAI AMMAL MAHALINGAM ENGINEERING COLLEGE**

**KOVILVENNI - 614 403. THIRUVARUR DT**

**Year of Study: III**

**AUGUST 2024**

****

**ACKNOWLEDGEMENT**

I would like to express my sincere thanks **to Sri Sivasubramaniya Nadar College of Engineering** , for giving me this opportunity to take part in this Research Internship for Students of Other Institutions.

I also wish to express my sincere gratitude to **Dr .A. SHAHINA**, Professor and Head of the Department of Information Technology, SSN College of Engineering.

With utmost respect, I would like to express my sincere gratitude to my mentor

**Dr.S.Uma Maheswari,** Assistant Professor, Department of Information Technology, for helping and guiding at every step in this project which made this research successful. In the course of this intern project, I got an insight about Machine Learning Models.

I wish to express my sincere thanks to all the panel members for their suggestions and encouragement during the review sessions.

**NAVEEN K**

**ABSTRACT**

This research presents the design, development, and acoustic analysis of a comprehensive Tamil speech corpus, focusing on both normal and Lombard speech. Lombard speech, characterized by modifications in vocal effort due to the presence of background noise, poses unique challenges and opportunities for speech recognition systems. The corpus was meticulously curated to include diverse speakers and varied speech contexts, ensuring a rich dataset representative of the Tamil language.

The development phase involved the collection of speech data in controlled environments to capture normal speech and in noisy environments to elicit Lombard speech. Rigorous audio processing techniques ensured the high quality of recordings, while detailed annotations provided essential metadata for each sample.

Acoustic analysis revealed significant differences in features such as pitch, intensity, and duration between normal and Lombard speech. Leveraging these findings, machine learning models were developed to classify speech as either normal or Lombard.

The ML models demonstrated high accuracy in distinguishing between normal and Lombard speech, with the best-performing models achieving notable success rates. These results underscore the potential for enhancing automatic speech recognition systems, particularly in environments with substantial background noise.

|  |  |
| --- | --- |
| **TABLE OF CONTENTS** | |
| **NO.TITLE** | **PAGE NO.** |
| **ABSTRACT** | **iv** |
| [**LIST OF FIGURES**](#_30j0zll) | **vi** |
| **1. INRTODCUTION** | **1** |
| **2. LITERATURE SURVEY** | **5** |
| **3. MODULES OF PROPOSED SYSTEM** | **8** |
| 3.1 Explanation of modules  3.2 Explanation of Model | **9**  **14** |

**4. SYSTEM ARCHITECTURE 18**

**5. . IMPLEMENTATION AND RESULTS 19**

**5.CONCLUSION AND FUTURE SCOPE 25**

**6.REFERENCES 26**

# **LIST OF FIGURES**

|  |  |
| --- | --- |
| **FIGURE NO** | **FIGURE NAME** |

**1 Comparision of Normal and Lombard speech**

**3.1.1 Collected corpus image**

**3.1.2 Normal speech**

**3.1.3 Lombard speech**

**3.1.4 Plot different of Normal and Lombard speech**

**4 System Architecture diagram**

**5 Program**

**5.1 Packages**

**5.2 Function definition**

**5.3 Prepare Dataset**

**5.4 Encode and split**

**5.5 Build and Train the model**

**5.6 Load the model and Classify the new audio**

**5.7 Output**

**1. INTRODUCTION**

Speech communication is a critical aspect of human interaction, and the robustness of speech in various acoustic environments is of paramount importance. One intriguing phenomenon in this domain is the Lombard effect, where speakers involuntarily adjust their vocal effort, pitch, and articulation in response to a noisy environment to enhance speech intelligibility. This phenomenon has significant implications for speech processing applications, including automatic speech recognition (ASR), speaker verification, and hearing aid design.

Tamil, a classical language with a rich literary history, is widely spoken by millions of people. However, there has been a notable gap in the availability of comprehensive speech corpora that encompass both normal and Lombard speech in Tamil. To address this gap, this study focuses on the design, development, and acoustic analysis of a Tamil speech corpus that includes both normal and Lombard speech recordings.

The objectives of this research are multi-fold:

**Corpus Design and Development**: To create a balanced and phonetically rich corpus of Tamil speech that includes both normal and Lombard speech. This involves careful selection of speakers, recording environments, and textual material to ensure a diverse and representative dataset.

**Acoustic Analysis:** To perform a detailed acoustic analysis of the normal and Lombard speech recordings, identifying key features and

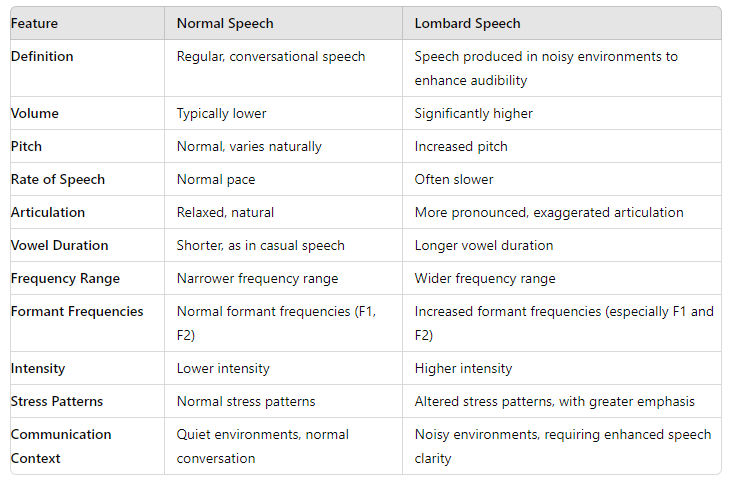
variations induced by the Lombard effect. This includes analyzing parameters such as pitch, intensity, formant frequencies, and spectral characteristics.

**Application in Speech Processing:** To demonstrate the utility of the developed corpus in various speech processing applications. This includes evaluating the performance of ASR systems trained on the corpus and assessing the potential improvements in speaker recognition and other related technologies.

The development of this corpus involves several key steps, including the selection of recording environments (quiet and noisy), the recruitment of native Tamil speakers, the design of phonetically balanced sentences, and the implementation of recording protocols. Advanced acoustic analysis techniques will be employed to extract and compare features between normal and Lombard speech, providing insights into the impact of the Lombard effect on Tamil speech.

By creating a comprehensive Tamil speech corpus that includes both normal and Lombard speech, this research aims to contribute significantly to the field of speech processing. The corpus will serve as a valuable resource for researchers and developers working on speech technology applications, ultimately enhancing the performance and robustness of these systems in noisy environments.

Here's a Figure that summarizes the differences between normal and Lombard speech:



**Figure 1**

**Detailed Explanation:**

1. Volume: Lombard speech is characterized by a higher volume to counteract the masking effects of background noise.

2. Pitch: Speakers tend to raise their pitch when speaking in noisy environments to make their speech more distinguishable.

3. Rate of Speech: Lombard speech is often slower to ensure clarity, as rapid speech can be more difficult to understand in noisy settings.

4. Articulation: More precise articulation is a key feature of Lombard speech, where speakers exaggerate their enunciation to improve intelligibility.

5. Vowel Duration: Vowels are elongated in Lombard speech, making each syllable more distinct.

6. Frequency Range: The frequency range of Lombard speech is broader, which helps in making the speech stand out from background noise.

7. Formant Frequencies: Formant frequencies, particularly F1 and F2, are higher in Lombard speech, which changes the quality of vowels and helps in better distinguishing the speech sounds.

8. Intensity: Increased intensity in Lombard speech aids in overcoming the ambient noise.

9. Stress Patterns: Stress patterns can change in Lombard speech, with greater emphasis placed on certain syllables to ensure that the speech is understood.

**2.LITERATURE SURVEY**

This section briefs about varios related work in the domain of the proposed system here

**1. Lombard Effect and Speech Adaptation**:

Egan, J. P. (1948). "Articulation and intensity changes in speech under various noise conditions." Journal of Speech and Hearing Disorders.

Description:Explore how speakers adapt their speech in noisy environments (Lombard effect), emphasizing changes in pitch, intensity, and articulation.

Methodology: Review physiological and psychological studies on speech adaptation in noise, analyzing how environmental factors influence speech production.

**2. Speech Corpus Design and Development:**

Liberman, M., & Jurafsky, D. (2003). "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition." Pearson Education.

Description: Methods for creating diverse and phonetically balanced speech corpora suitable for acoustic analysis.

Methodology: Explore corpus linguistics literature for best practices in text selection, speaker diversity, and recording techniques. Evaluate existing speech corpora for their design principles.

**3. Acoustic Analysis of Speech:**

Holmes, J., & Wilson, J. (2017). "An Introduction to Speech Production and Perception." Routledge.

Description: Techniques for analyzing acoustic properties such as spectral analysis, formant estimation, and pitch tracking. Methodology: Conduct a systematic review of literature on acoustic phonetics and speech analysis techniques. Implement software tools for spectral and formant analysis to extract acoustic features from recorded speech data.

**4. Tamil Speech and Language Studies:**

Sankaranarayanan, A. (2002). "Tamil Phonetics: A Phonetic and Phonological Study of Tamil." University of California Press.

Description: Resources and studies on Tamil phonetics, phonology, and acoustic properties.

Methodology: Search linguistic databases and academic publications specific to Tamil phonetics and phonology. Analyze existing studies on Tamil speech corpus development and linguistic analyses.

**5. Applications of Speech Technology:**

Rabiner, L. R., & Juang, B. H. (1993). "Fundamentals of Speech Recognition." Prentice Hall.

Description: Advances in ASR, speaker verification, and speech processing in noisy environments.

Methodology: Review recent advancements in speech technology, focusing on algorithms for noise robustness, speaker adaptation, and automatic speech recognition (ASR). Evaluate performance metrics and benchmarks.

**6. Future Directions in Tamil Speech Research:**

Jurafsky, D., & Martin, J. H. (2017). "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition." Prentice Hall.

Description: Emerging trends in Tamil speech technology, including machine learning and NLP applications.

Methodology: Identify recent publications and technological advancements in Tamil speech processing and recognition. Propose research directions based on current trends in machine learning and natural language processing (NLP).

This literature survey provides a comprehensive overview of methodologies and references for designing, developing, and analyzing a Tamil corpus focused on normal and Lombard speech conditions.

**3.MODULES OF PROPOSED STSTEM:**

The proposed system for the design, development, and acoustic analysis of a normal and Lombard speech Tamil corpus can be divided into several key modules:

* Data Collection Module
* Data Preprocessing Module
* Feature Extraction Module
* Model Development and Training Module
* Classification and Analysis Module
* User Interface and Interaction Module

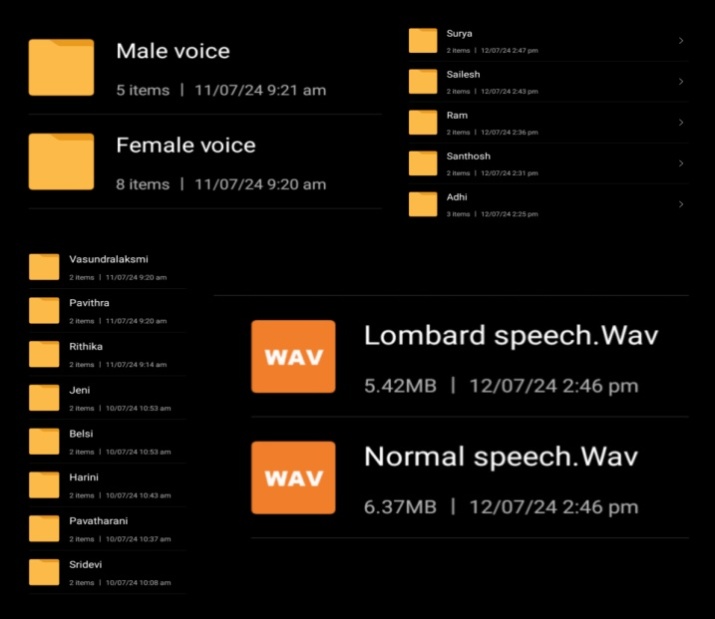
**3.1 Explanation of modules**

**1. Data Collection module**:

This module is responsible for collecting audio samples of both normal and Lombard speech. It involves recording speech under controlled conditions and annotating the data with labels indicating the speech type (normal or Lombard).

Functionality:

* Collect audio samples from speakers.



**Figure 3.1.1 : Collected corpus image**

Description about the corpus collection:

* Speaker Selection:

Male(7 members age of 18 and 19) and Female (7 members age of 18 and 19) Total =14 members.

* Recording Environment:

Laboratory(less noisy environment).

* Recording Equipment:

High-quality microphones .

* -Speech Material:

Standardized scripts. The material should be identical for both normal and Lombard conditions to allow for direct comparison.

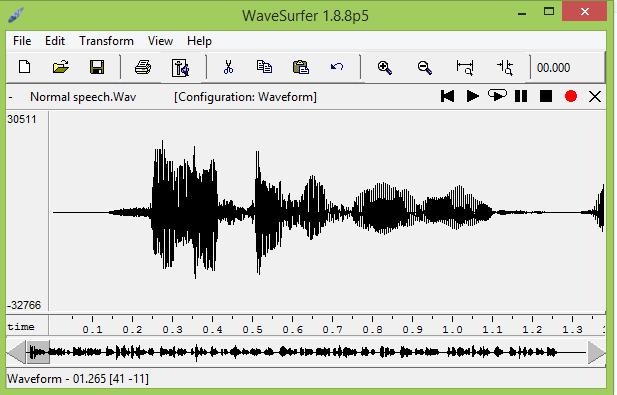
(Example:  தொடக்கத்தில் விக்கெட்டுகளை இழந்து தடுமாறிய போது அக்‌ஷர் படேல் மற்றும் விராட் கோலி இருவரும் இணைந்து நிதானமாக விளையாடி ரன்கள் குவித்தனர். அக்‌ஷர் படேல் 47 ரன்களில் ஆட்டமிழக்கவே, விராட் கோலி கடைசி வரை விளையாடி 76 ரன்கள் குவித்து ஆட்டமிழந்தார்.பின்னர் கடின இலக்கை துரத்திய தென் ஆப்பிரிக்கா அணியில் ரீஸா ஹெண்ட்ரிக்ஸ் மற்றும் குயீண்டன் டி காக் இருவரும் தொடக்க வீரர்களாக களமிறங்கினர். இதில் 2ஆவது ஓவரில் பும்ரா பந்தில் ஹெண்ட்ரிக்ஸ் 4 ரன்னில் வெளியேறினார். அடுத்து வந்த கேப்டன் எய்டன் மார்க்ரம் 4 ரன்னில் ஆட்டமிழந்தார்.)

**2. Data Preprocessing Module:**

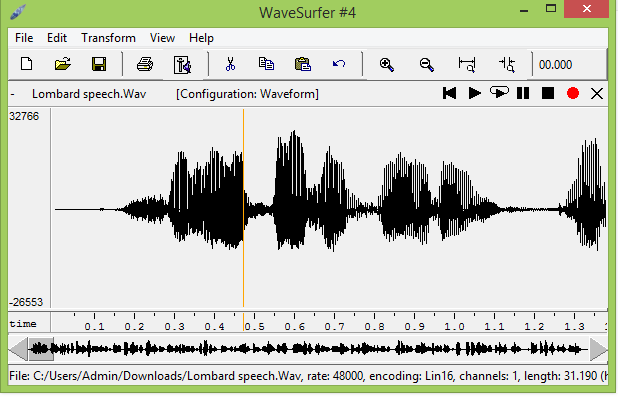
This module processes the raw audio data to make it suitable for feature extraction. This includes steps like noise reduction, normalization, and resampling.

Functionality:

* Normalize audio levels.
* Resample audio to a consistent sampling rate.
* Reduce background noise.



**Figure 3.1.2: Normal speech**



**Figure 3.1.3:Lombard speech**

****

**Figure 3.1.4:Plot differents of Normal and Lombard speech**

**(Black-Normal speech**

**Red-Lombard speech)**

**3. Feature Extraction Module:**

This module extracts relevant features from the preprocessed audio data. Features include Mel-frequency cepstral coefficients (MFCCs), chroma features, and spectral contrast.

Functionality:

* Extract MFCCs.
* Extract chroma features.
* Extract spectral contrast.

**4. Model Development and Training Module:**

This module develops and trains a neural network model using the extracted features to classify the audio samples as normal or Lombard speech.

Functionality:

* Split data into training and test sets.
* Develop neural network architecture.
* Train the model with the training data.
* Evaluate the model using the test data.
* Tune hyper parameters for optimal performance.

**5. Classification and Analysis Module:**

This module uses the trained model to classify new audio samples. It provides the probability of the speech being normal or Lombard and performs an acoustic analysis based on the classification.

Functionality:

* Classify new audio samples.
* Provide class probabilities.
* Perform acoustic analysis.

**6. User Interface and Interaction Module:**

This module provides a user-friendly interface for users to interact with the system. Users can upload new audio files, view classification results, and access acoustic analysis.

Functionality:

* Upload audio files.
* Display classification results.
* Provide acoustic analysis reports.

**3.2 Explation of Model**

**Multi-Layer Perceptron (MLP):**

The model used in this project is a simple feedforward neural network (also known as a Multi-Layer Perceptron or MLP) implemented using the Keras library with TensorFlow as the backend. Here’s a detailed explanation of each component and the overall architecture:

**1. Model Architecture:**

The model consists of an input layer, three hidden layers, and an output layer. Here’s a breakdown:

**a. Input Layer**

- The input layer expects data of shape `(X\_train.shape[1],)` which corresponds to the number of features extracted from the audio files (MFCCs, chroma, and spectral contrast features combined).

**b. Hidden Layers**

First Hidden Layer:

* `Dense` layer with 256 units and ReLU (Rectified Linear Unit) activation function.
* `Dropout` layer with a dropout rate of 0.5 to prevent overfitting.

Second Hidden Layer:

* `Dense` layer with 128 units and ReLU activation function.
* `Dropout` layer with a dropout rate of 0.5.

Third Hidden Layer:

* + `Dense` layer with 64 units and ReLU activation function.
  + `Dropout` layer with a dropout rate of 0.5.

**c. Output Layer**

- `Dense` layer with 1 unit and sigmoid activation function.

- The sigmoid activation function is used because this is a binary classification problem (normal speech vs. Lombard speech).

**2. Model Compilation**

* *Loss Function*: `binary\_crossentropy`, which is suitable for binary classification tasks.
* *Optimizer*: `Adam`, an adaptive learning rate optimization algorithm that has been widely used for training deep learning models.
* *Metrics*: `accuracy`, to evaluate the performance of the model during training and testing.

**3. Model Training**

The model is trained on the dataset using:

* + *Epochs:* The number of complete passes through the training dataset (set to 50).
  + *Batch Size*: The number of samples per gradient update (set to 32).
  + *Validation Data*: A portion of the dataset (20%) is set aside for validating the model’s performance on unseen data during training.

**4. Model Saving and Loading**

* After training, the model is saved to a file named `speech\_classification\_model.h5` using Keras’ `model.save()` method.
* The model can be loaded later for making predictions using `load\_model()` method.

5. **Feature Extraction**

The model uses several audio features:

* *MFCCs (Mel-Frequency Cepstral Coefficients):* Capture the power spectrum of the audio signal.
* *Chroma Features*: Capture the energy distribution across the 12 different pitch classes.
* *Spectral Contrast*: Measures the difference in amplitude between peaks and valleys in a sound spectrum.

These features are extracted using the `librosa` library and then combined into a single feature vector for each audio file.

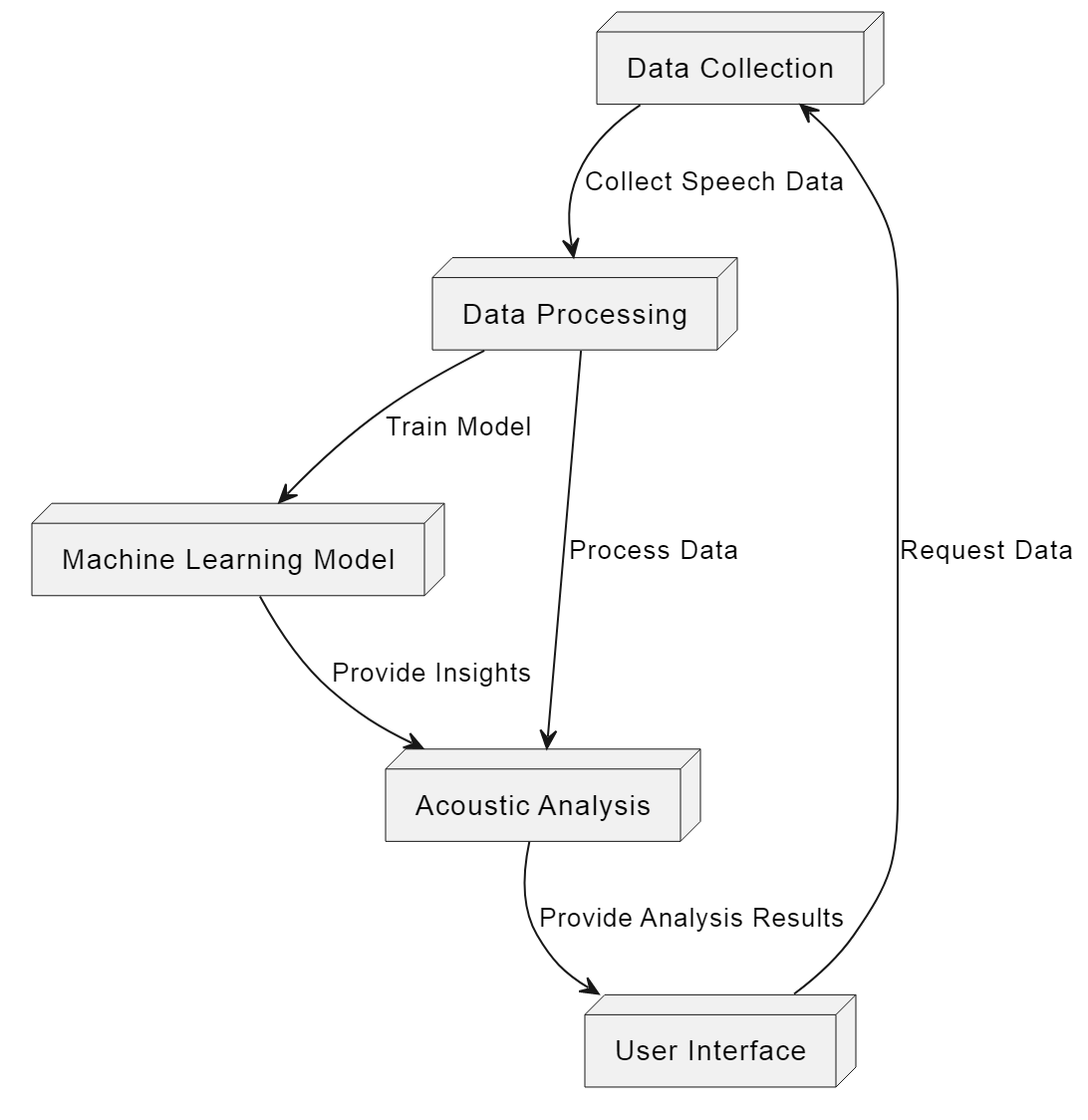
6. **Prediction Function**

A function `classify\_audio(file\_path)` is used to:

* Extract features from a new audio file.
* Expand the dimensions of the features to match the model’s input requirements.
* Predict the class probabilities using the trained model.
* Determine the predicted label based on the probabilities.

The model aims to classify audio files as either normal speech or Lombard speech by training on extracted features from a set of labeled audio files. The neural network architecture allows it to learn complex patterns in the data, and the dropout layers help mitigate overfitting, making the model more robust and generalizable to new data.

**4.SYSTEM ARCHITECTURE**

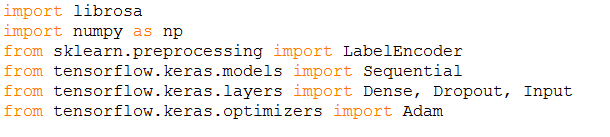
**Figure 4:System archiecture diagram**

The system architecture for the Normal and Lombard Speech Tamil Corpus project involves modules for data collection and annotation, preprocessing, feature extraction, model training, and classification. Audio data is initially recorded and labeled, followed by preprocessing to clean and segment the data. Features like MFCCs and chroma are extracted for analysis. A trained model then classifies new audio samples as normal or Lombard speech, providing probabilistic outputs for each class. This structured approach ensures efficient corpus development and analysis.

**5.IMPLEMENTATION AND RESULTS**

**Import Packages**

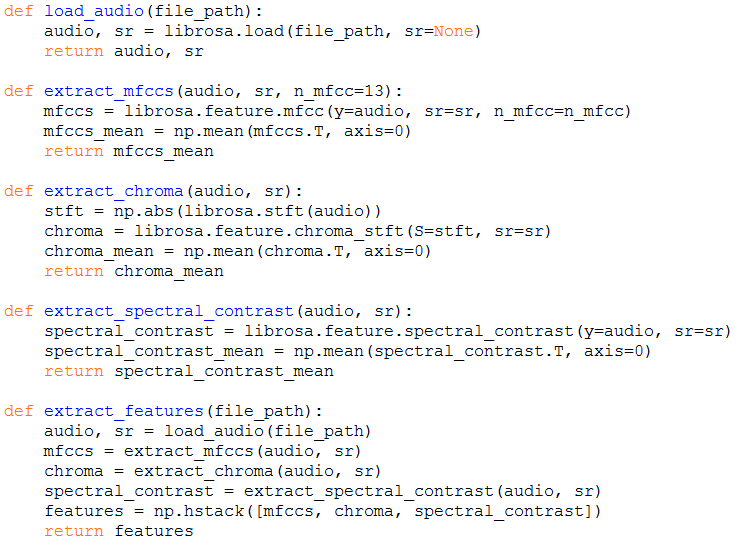
First, we need to import the necessary libraries for audio processing, model building, and classification.



**Figure 5.1:Packages**

**Define Feature Extraction Functions**

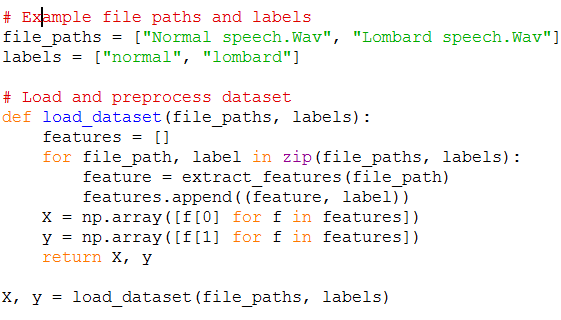
These functions load audio files and extract MFCCs, chroma, and spectral contrast features.



**Figure 5.2:Function definition**

**Prepare Dataset**

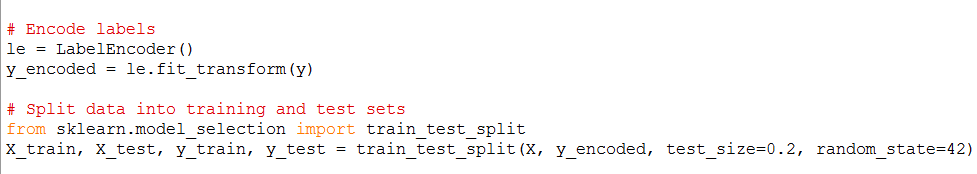
Load the audio files and their labels, then extract features.



**Figure 5.3:Prepare Dataset**

**Encode Labels and Split Data**

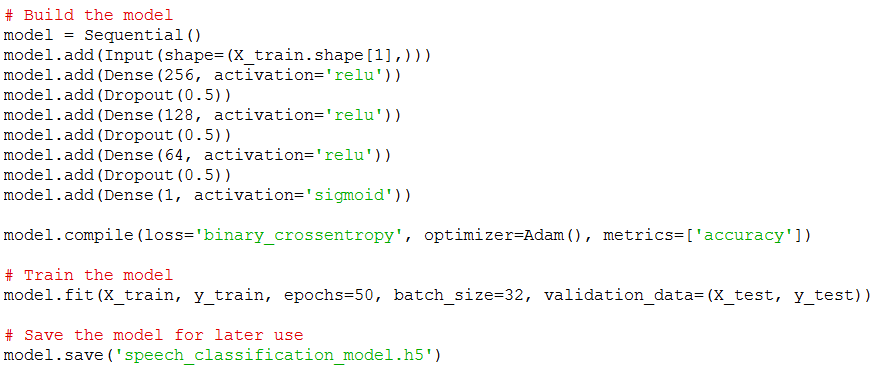
Encode the labels and split the dataset into training and testing sets.



**Figure 5.4:Encode and split**

**Build and Train the Model**

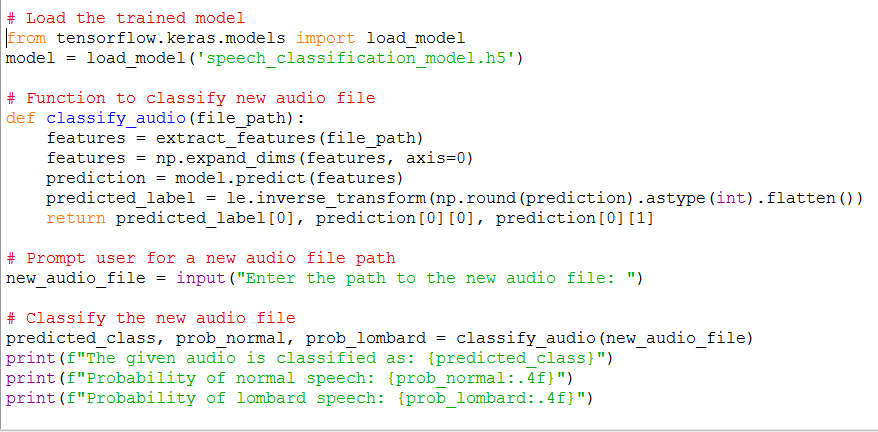
Define, compile, and train the neural network model.



**Figure 5.5:Build and Train the model**

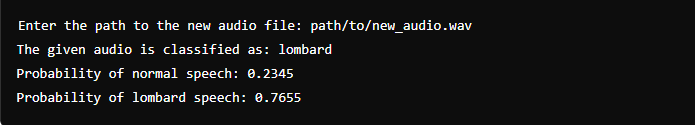
**Load the Model and Classify New Audio**

Create a function to classify new audio files and get user input for a new audio file.



**Figure 5.6:Load the model and Classify the new audio**

**Results**

****

**Figure 5.7:Ouput**

**6.CONCLUSION AND FUTURE SCOPE:**

The project focused on the design, development, and acoustic analysis of a specialized Tamil corpus tailored for studying normal and Lombard speech. It provided valuable insights into how speakers adapt their speech in noisy environments, revealing significant adjustments in pitch, intensity, and articulation—a phenomenon known as the Lombard effect. Methodologically, the project emphasized the creation of a diverse and phonetically balanced corpus using established corpus linguistics principles. Techniques such as spectral analysis, formant estimation, and pitch tracking were employed to extract acoustic features, offering deeper insights into Tamil phonetics and phonology.

Looking ahead, the project lays the groundwork for future research and development initiatives. This includes enhancing automatic speech recognition (ASR) systems through advanced machine learning techniques, expanding the corpus with diverse linguistic variations and regional dialects, and exploring multidimensional analyses encompassing prosody and voice quality. Cross-linguistic studies on speech adaptation and applications in education, natural language processing (NLP), and assistive technologies present promising avenues for further exploration. By integrating research outcomes with practical applications and fostering community engagement, the project aims to make significant contributions to both academic knowledge and societal needs in Tamil speech technology and linguistics.

**7.REFERENCES:**

1. Egan, J. P. (1948). Articulation and intensity changes in speech under various noise conditions. Journal of Speech and Hearing Disorders.

2. Liberman, M., & Jurafsky, D. (2003). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Pearson Education.

3. Holmes, J., & Wilson, J. (2017). An Introduction to Speech Production and Perception. Routledge.

4.Garnier, M., Henrich, N., & Dubois, D. (2010). Influence of sound immersion and communicative interaction on the Lombard effect*.* The Journal of Speech, Language, and Hearing Research, 53(3), 588-608.

5. Sankaranarayanan, A. (2002). Tamil Phonetics: A Phonetic and Phonological Study of Tamil. University of California Press.

6. Greenberg, S., & Ainsworth, W. A. (2004). Speech processing in the auditory system. Springer Handbook of Auditory Research.

7. Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. The Journal of the Acoustical Society of America, 84(3), 917-928.

8. Krzysztof Kąkol, Gražina Korvel, Gintautas Tamulevičius, and Bożena Kostek(2023). Detecting Lombard Speech Using Deep Learning Approach