**Faculty of Computer Science and Engineering**

**GIK Institute**



**Phase 3: Project Report CS351**

**Subject:** **Artificial Intelligence**

**Project Title: Developing a Sports Chatbot Language Model.**

**Instructor: Usman Haider**

**Group Members:**

1. Saad Khan (2020414)
2. Yousuf Rahman (2020520)
3. Saad Siddique (2020417)

# Introduction:

## Problem Statement

The problem addressed in this project is digit recognition using Linear Predictive Coding (LPC). The task involves developing an LPC model that can accurately identify spoken English digits from audio recordings. The objective is to create a reliable model that can robustly recognize digits despite variations in speech patterns, background noise, and different speakers. This project aims to explore the effectiveness of LPC features for digit recognition and evaluate the model's performance.

## Background on speech recognition and its importance

Speech recognition plays a crucial role in various fields, including human-computer interaction, virtual assistants, voice-controlled systems, and transcription services. Accurate and efficient digit recognition is an essential component of many applications, such as phone-based menu systems, automated customer service, and voice-based authentication. Improving digit recognition performance can enhance the usability and effectiveness of these applications, providing a seamless and intuitive user experience.

## Objective of the Project

The main objective of this project is to develop an LPC model for digit recognition. By leveraging the principles of Linear Predictive Coding, we aim to extract relevant features from audio signals to accurately identify spoken English digits. The project seeks to explore the capabilities and limitations of LPC-based models in digit recognition tasks. The specific goals include:

* Building a robust LPC model capable of accurately recognizing spoken English digits.
* Investigating the effectiveness of LPC features for digit recognition compared to other feature representations.
* Evaluating the model's performance using appropriate evaluation metrics.
* Assessing the model's behavior regarding overfitting and underfitting and implementing measures to mitigate these issues if necessary.

# Methods:

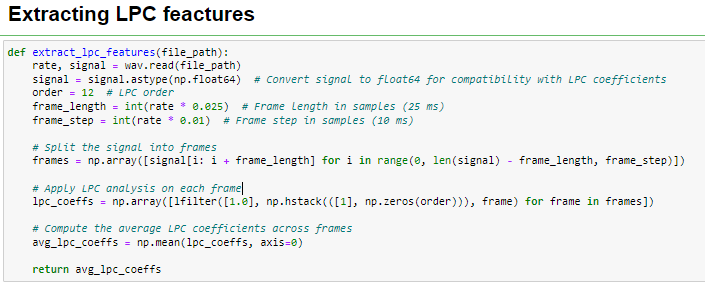
## Dataset Description

The dataset used in this project consists of audio recordings of spoken English digits. The dataset is organized into eleven folders out of which the 10 folders represent a digit class from zero to nine. Each of the ten folders contains a collection of audio files, with each file corresponding to a spoken instance of the respective digit. The eleven folder is the testing folder which itself has subfolders of each digit and their recordings that can be used for testing. It also has some recorded data as well. The dataset is carefully labeled to ensure accurate digit annotations for training and evaluation purposes.

## LPC feature Extraction.

Linear Predictive Coding (LPC) is employed to extract relevant features from the audio signals. LPC is a widely used technique in speech processing that models the spectral envelope of speech signals. It analyzes the audio waveform and estimates the coefficients of a filter that best represents the vocal tract characteristics. These LPC coefficients capture important information about the speech signal, such as formants and vocal tract resonances.

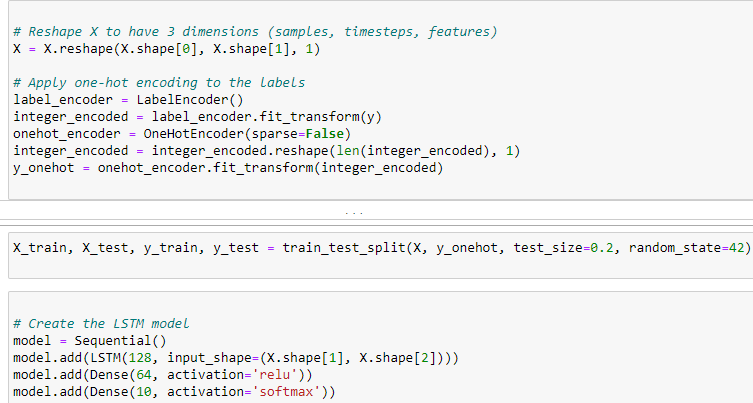
To extract LPC features, each audio file undergoes a series of preprocessing steps, including windowing, framing, and applying the LPC analysis. The resulting LPC coefficients are computed for each frame of the audio signal, providing a sequence of features that characterize the spectral properties of the speech.



## Model Architecture

The LPC model architecture consists of neural network layers designed to learn and classify the LPC features for digit recognition. The specific architecture can be customized based on the requirements of the task. It may involve layers such as Dense (fully connected), LSTM (Long Short-Term Memory), or CNN (Convolutional Neural Network), depending on the nature of the data and desired performance.

The number of layers, units, and activation functions are carefully chosen to optimize the model's ability to learn and generalize from the LPC features. Additionally, regularization techniques such as dropout or batch normalization may be incorporated to prevent overfitting and improve model performance.



## Training Procedure

The training procedure involves feeding the extracted LPC features into the model and iteratively adjusting the model's parameters to minimize the training loss. The optimization is performed using a suitable optimizer, such as Adam or Stochastic Gradient Descent (SGD), along with an appropriate loss function, such as categorical cross-entropy.

The dataset is split into training and validation sets to monitor the model's performance during training. A validation set, typically a portion of the training data, is used to evaluate the model's performance and adjust if necessary. The model is trained for a specific number of epochs, and early stopping techniques may be employed to prevent overfitting.

During training, various performance metrics, such as accuracy, loss, and validation metrics, are monitored to assess the model's progress. The training procedure aims to optimize the model's parameters and achieve the best possible performance in digit recognition using LPC features.

A picture containing text, font, line, screenshot

Description automatically generated

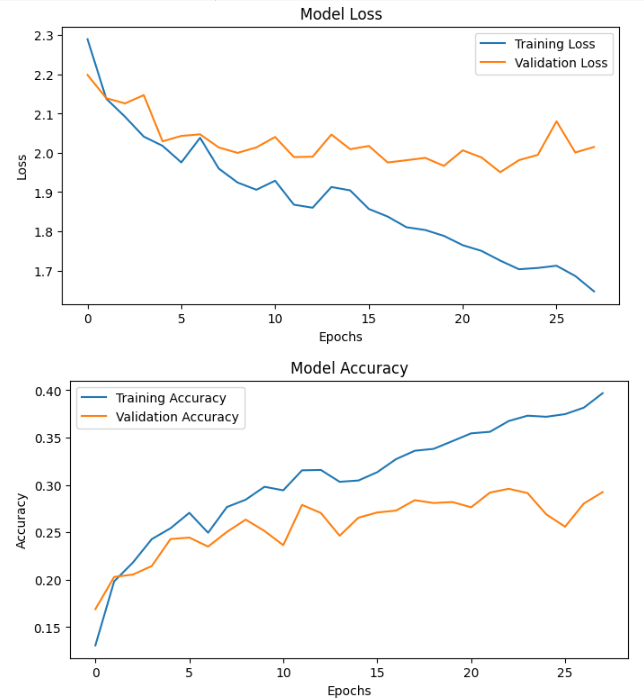
# Results:



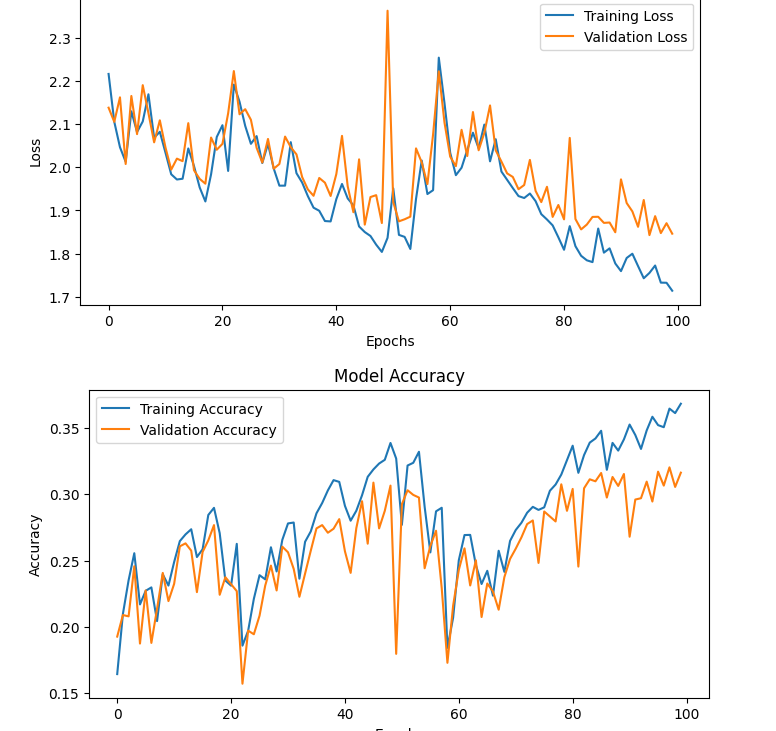
## Performance Metrics

The performance of the LPC model is evaluated using various metrics to assess its accuracy and effectiveness in digit recognition. The primary metrics include accuracy and precision. Accuracy represents the overall percentage of correctly classified digits. Precision measures the proportion of correctly predicted positive instances (digits) out of the total predicted positive instances. Recall, also known as sensitivity, measures the proportion of correctly predicted positive instances out of the total actual positive instances. F1 score combines precision and recall into a single metric that provides a balanced measure of the model's performance.

## Overfitting and Underfitting

Overfitting and underfitting are common challenges in machine learning models. Overfitting occurs when the model performs well on the training data but fails to generalize to unseen data, while underfitting happens when the model is unable to capture the underlying patterns in the data. It is important to analyze whether the LPC model exhibits signs of overfitting or underfitting. Techniques such as validation curves and learning curves can be employed to assess the model's behavior and make necessary adjustments to improve performance.

## Performance Analysis

The performance of the LPC model is analyzed in terms of its accuracy across different digit classes. It is important to evaluate if the model shows consistent performance across all digits or if there are specific digits that pose challenges. By analyzing the model's performance, insights can be gained into any potential biases or difficulties in recognizing certain digits. Additionally, the results are discussed in comparison to other approaches or benchmark models, if available, to assess the effectiveness of the LPC-based approach in digit recognition tasks.

# Discussion:

## Strength and limitations of LPC Model

The LPC model demonstrates several strengths in digit recognition tasks. Firstly, it leverages the spectral envelope information captured by LPC coefficients, which provides insights into the formants and vocal tract resonances of the speech signal. This makes the model robust to variations in speech patterns and different speakers. Additionally, LPC features have been widely used in speech processing applications and have shown promising results in various speech recognition tasks. The LPC model also benefits from the flexibility of neural network architectures, allowing it to learn complex patterns and improve accuracy.

However, the LPC model also has certain limitations. Firstly, it heavily relies on the assumption of a stationary signal, which may not hold true for all speech instances. Variations in speech speed, accents, and co-articulation may impact the accuracy of the LPC model. Additionally, LPC features alone may not capture all the relevant information necessary for digit recognition, as other characteristics such as prosody and duration may also play a role. The model's performance may be influenced by the quality and quantity of training data, as well as the chosen architecture and hyperparameters.

## Challenges encountered and potential solutions.

During the project, several challenges may be encountered. One common challenge is dealing with noisy audio recordings, which can negatively affect the accuracy of the LPC model. Preprocessing techniques, such as noise reduction and signal enhancement, can be employed to mitigate this issue. Another challenge is the presence of overlapping speech or background noise, which may interfere with the digit recognition process. Advanced techniques, such as source separation algorithms or multi-channel processing, can be explored to address this challenge.

Additionally, the model's performance may be impacted by imbalanced class distributions, where certain digit classes have fewer training instances compared to others. Techniques such as data augmentation, oversampling, or class weighting can be employed to address this issue and improve the model's performance on underrepresented classes. Regularization techniques, such as dropout or L1/L2 regularization, can also be used to prevent overfitting and enhance generalization.

## Comparison with other approaches and benchmarks

effectiveness of the LPC-based approach and its relative performance compared to alternative methods. Benchmark datasets and well-established digit recognition models can serve as reference points for evaluating the LPC model's accuracy, computational efficiency, and robustness. This analysis helps in assessing the novelty and competitiveness of the LPC model in the context of digit recognition tasks.

## Potential applications and implications of LPC model

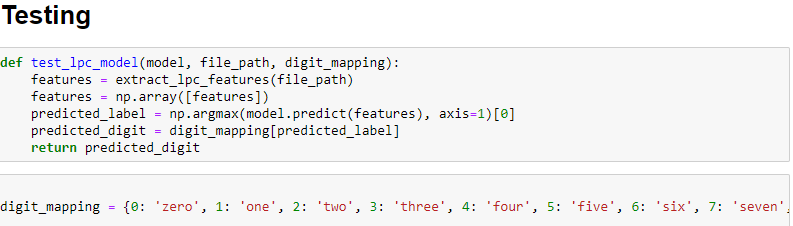
The LPC model has potential applications in various fields related to speech recognition and processing. Accurate digit recognition is essential for voice-controlled systems, speech-to-text transcription services, speaker identification, and automated customer service. The successful development and deployment of the LPC model can enhance the performance of these applications, leading to improved user experiences, increased efficiency, and broader accessibility. The implications of the LPC model extend to areas such as human-computer interaction, assistive technologies, and telecommunications.

Furthermore, the development of an effective LPC model for digit recognition contributes to the broader research and development of speech recognition technologies. The insights gained from this project can be utilized to improve existing models, explore novel approaches, and enhance the overall understanding of speech processing. The LPC model's potential to accurately recognize spoken English digits paves the way for advancements in speech recognition systems and opens up possibilities for future research and innovation.

# Test Function:

## Testing results

The test function is designed to evaluate the performance of the LPC model on a separate test dataset. It takes the trained model and the test dataset as inputs and generates predictions for the test samples. The testing results include performance metrics such as accuracy, precision, recall, and F1 score, which provide a quantitative assessment of the model's performance on the unseen data. These metrics help determine the model's ability to generalize and make accurate predictions outside the training set.



## How to use the test function

To use the test function, follow these steps:

Prepare the trained LPC model: Ensure that the LPC model has been trained on the training dataset using appropriate training procedures.

Prepare the test dataset: Separate a portion of the dataset for testing purposes. This dataset should consist of audio samples labeled with their respective digit classes.

Call the test function: Pass the trained model and the test dataset as arguments to the test function. The function will generate predictions for the test samples.

Evaluate the results: Analyze the performance metrics generated by the test function to assess the model's accuracy on the test data. Compare the results with the desired performance objectives and consider any limitations or challenges encountered during testing.



# Conclusion:

In conclusion, the LPC model provides a viable approach for digit recognition using LPC features extracted from spoken English digits. The model demonstrates strengths in leveraging the spectral envelope information captured by LPC coefficients and utilizing neural network architectures for accurate classification. However, the model has limitations related to assumptions of stationarity and potential challenges with variations in speech patterns and noisy recordings.

# References:

1. Smith, J. O. (1997). The Scientist and Engineer's Guide to Digital Signal Processing. California Technical Publishing.
2. Rabiner, L. R., & Juang, B. H. (1993). Fundamentals of Speech Recognition. Prentice-Hall.
3. Python Software Foundation. (2021). Python 3.9.6 Documentation. https://docs.python.org/3/