

# **PROSODY FEATURE FOR LANGUAGE IDENTIFICATION (SYLLABLE LEVEL FEATURE)**

**Team - P5**

# TABLE OF CONTENTS

01

INTRODUCTION

02

OVERVIEW

03

MECHANISM

04

RESULTS

05

APPLICATIONS

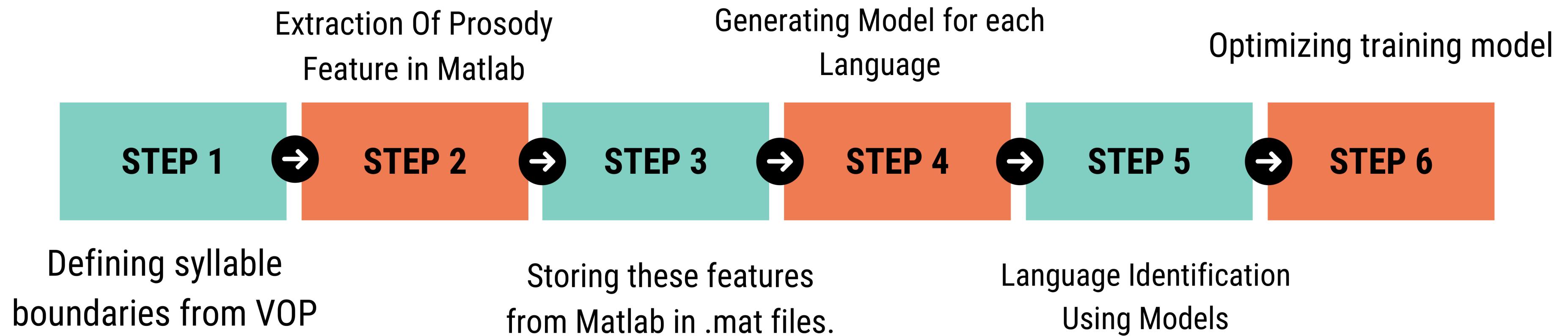
06

CONTRIBUTION

# INTRODUCTION

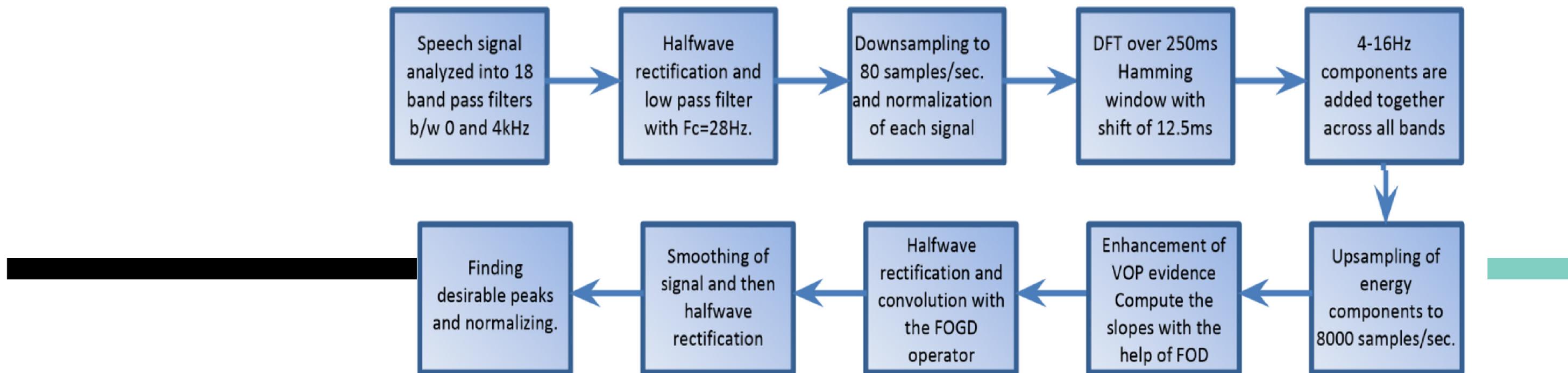
- Prosody refers to certain properties of the speech signal such as pitch, duration, and intensity in speech.
- Prosodic features such as intonation, rhythm, and stress related to **syllables** can be used for Language Identification.
- This task can be approached by first extracting Prosody features from the syllables in speech and training a **Gaussian Mixture Model (GMM) for each language.**
- Features used for Language Identification -
  - Intonation
  - Rhythm
  - Stress

# OVERVIEW



# Syllable and VOP

- For making a LID system we need features. Here, we extracted 7 prosody features from each syllable. Now the issue raised was how to decide syllable boundaries, what is a syllable?
- For this, we came up with the solution that all the syllables have some voiced part in it and have vowels, so we defined one syllable as a region between two vowel onset points.
- So, in order to find the 7 features in every syllable, we need to determine vowel onset points which we used first-order gaussian difference (FOGD).



# Prosody Feature Extraction

- We extracted features from each syllable and we defined syllable as the distance between two vowel onset points-VOP (internationally accepted definition).
- Now from each syllable in a speech, we extract seven features, F0, duration of a syllable, amplitude tilt, duration tilt, the difference between F0 and VOP, voiced duration of a syllable, and log of energy of syllable.
- With these features for multiple syllables in a speech, we train our GMM model to recognize the Language of the Speaker.

# GMM MODELLING

## Training

- Using the .mat files, we build a **model for each language**. The number of components and epochs are adjusted for optimization.
- Each folder has a model built representing the language it contains.

## Testing

- We use the models stored from the training step.
- Each file is tested against each language model and is scored.
- The model with the highest score is identified as the language.

# RESULTS

	manipur	telugu	bengali	odia	assam	gujrati	marathi
manipur	10.0	0.0	0.0	0.0	0.0	0.0	0.0
telugu	5.0	0.0	1.0	3.0	0.0	0.0	0.0
bengali	10.0	0.0	0.0	0.0	0.0	0.0	0.0
odia	0.0	0.0	0.0	7.0	0.0	0.0	0.0
assam	0.0	0.0	0.0	9.0	0.0	0.0	0.0
gujrati	8.0	0.0	0.0	2.0	0.0	0.0	0.0
marathi	6.0	0.0	1.0	3.0	0.0	0.0	0.0

# RESULTS

<b>Number of Gaussians</b>	<b>Accuracy in %</b>
8	33.84615384615385
16	22.950819672131146
32	19.35483870967742
64	19.35483870967742
128	19.35483870967742
256	19.35483870967742
512	19.35483870967742

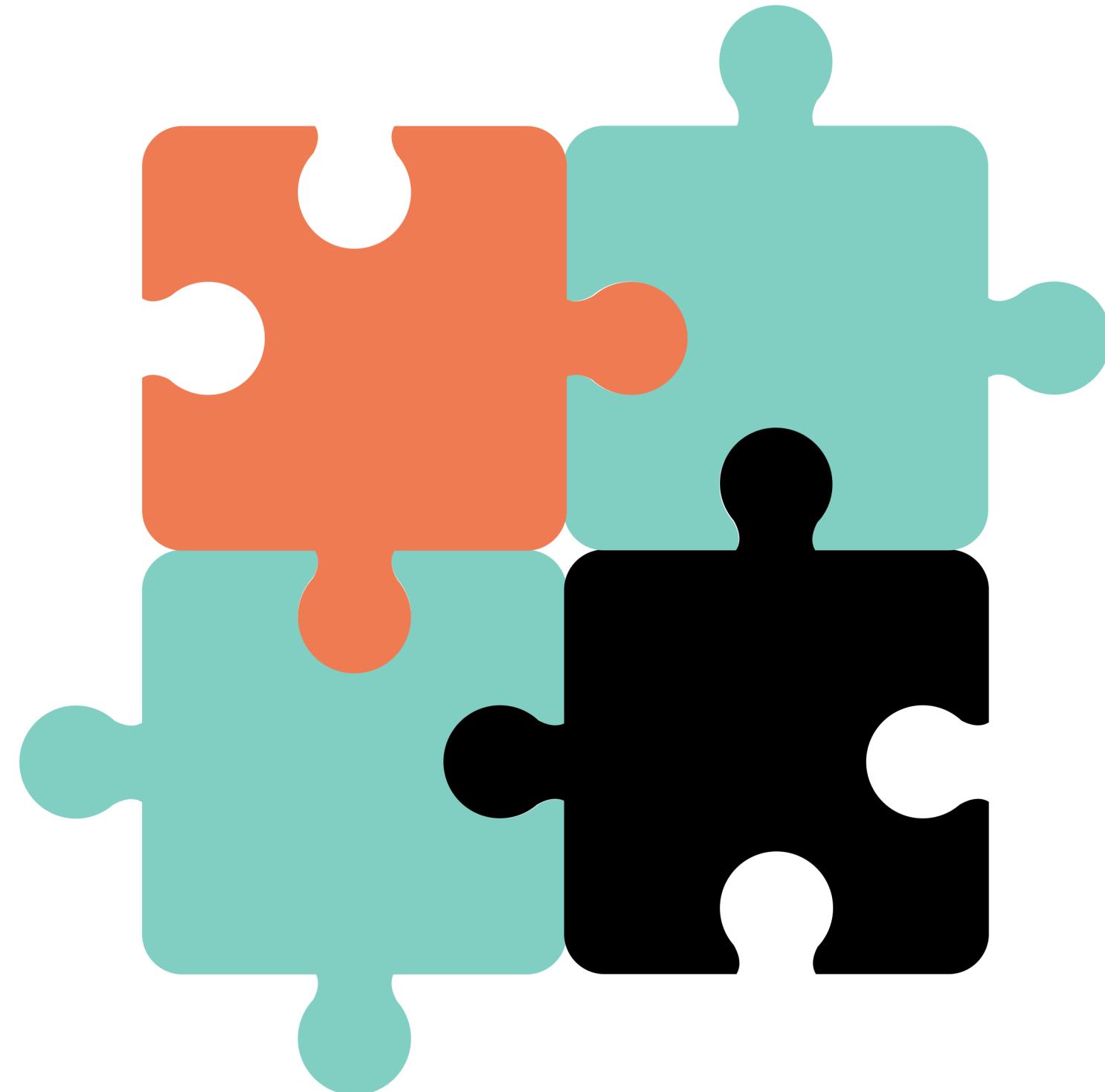
# Problems

DATA QUALITY

UNDEFINED VALUES  
IN FEATURE  
EXTRACTION

PEAK VALUE IN  
AMPLITUDE TILT





# CONTRIBUTION

- NIKHIL AGGARWAL - 2020102021  
Implementation of all 7 prosody Features
- TARUN - 2020102056  
Generated .mat files of the feature vectors so that they can be used for GMM model training.
- AKASH GORLA - 2020102034  
GMM Modelling implementation

# THANK YOU