**CT 478-Speech Technology**

Project Report

**Project Name:**

**Relative phase information for detecting human speech and spoofed speech**

**Guidance: Prof. Hemant Patil**

**Submitted by : Dhavalkumar Prajapati ( 201501188 )**



**Dhirubhai Ambani Institute of Information and Communication Technology**

**Gandhinagar,Gujarat,India**

**May 2018**

**TABLE OF CONTENTS**

[**Acknowledgement**](#_xvvzvztvdr0n) **3**

[**Terminology**](#_cdeym2qqxxn1) **4**

[**ABSTRACT**](#_e51bfi170mnz) **5**

**1.**[**Introduction**](#_7fod8wm0u4x0) **6**

[**2. Motivation**](#_visnhqzdvl4o) **6**

[**3. Application**](#_jeh0zby1et2h) **6**

[**4.System Architecture**](#_3qk6jyrx4ihw) **6**

[4.1 Overview of spoofing detection system using GMM](#_oerq34bzgm6h) 7

[4.2 Relative Phase Shift](#_sy1uxfkczrho) 10

[4.3 Group Delay](#_84ur4wq87i9v) 10

**5.**[**Novelty of My work**](#_wxwbxrc2vdiv) **11**

[**6. Algorithm Development**](#_ygmb4kaf63d8) **11**

[6.1 Algorithm Development of Relative phase shift](#_rxfy3u5e8zjh) 11

[6.2 Algorithm Development of Modified Group Delay cepstral coefficient](#_es0icpaip173) 12

[6.3 Algorithm Development of Group Delay](#_khxyulbiffdt) 12

[**7.pseudo code**](#_jvg4nbf9t1ly) **13**

[7.1 pseudo code for RPS](#_xkeokja5vyyp) 13

[7.2 pseudo code for Modified Group Delay](#_yr68477i6qlx) 13

[**8. Result and Analysis**](#_sls9xi9rd475) **14**

[8.1 Result and Analysis for Relative Phase shift](#_ey88f611d3sj) 14

[8.2 Result and Analysis for Group delay](#_983zvaa9ctk7) 15

[8.3 Result and Analysis for Modified Group Delay Cepstral Coefficient](#_dmxsryfx7qxw) 16

[8.4 comparison of result](#_propebfvle8s) 16

[**9.Limitations**](#_1ztznbwitenn) **17**

[**10.Future work**](#_xolkbtdx4bg) **17**

[**11.References**](#_nijp34hpyolp) 18

# **Acknowledgement**

The satisfaction and euphoria on the successful completion of any task would be incomplete without mentioning the people who made it possible whose constant guidance and encouragement crowned out effort with success.

We would like to express our heartfelt gratitude to our esteemed supervisor, **Prof. Hemant Patil** for his technical guidance, valuable suggestions, and encouragement throughout the experimental and theoretical study and in this course project. It has been our honor to work under his guidance, whose expertise and discernment were keys in the completion of this project.

We are grateful to the **DAIICT,** for giving us the opportunity to execute this project, which is an integral part of the curriculum in B.Tech programme **CT478** Speech Technology course at the DAIICT, Gandhinagar.

Many thanks to **Mr. Srinivas bhai** and our friends who are directly or indirectly helped us in our project work for their generous contribution towards enriching the quality of the work.

This acknowledgment would not be complete without expressing our sincere gratitude to **Speech Lab** for the help.

# Terminology

RPS : Relative Phase Shift

GMM : Gaussian Mixture Model

MGDCC : Modified Group Delay Cepstral Coefficient

MFCC : Mel frequency cepstral coefficient

GD : Group Delay

DET : Detection Error Trade off

# **ABSTRACT**

Recently, speaker verification technology has been used in many applications, such as a telephone banking, customer services,Authentication of gates, so security is very important and needs to prevent from spoofing attack.Any professional can make synthetic speech which is exactly same as natural speech, Anyone can use replay speech as spoof speech,which is more easy then other attacks. There are many different kinds of methods for detection of human speech and spoofed speech. However, in this project I am going to use relative phase information extracted from a Fourier spectrum is used to detect human and spoofed speech. Because original/natural phase information is almost entirely lost in spoofed speech,so in this project i will analysis the Relative phase shift of signal and extract RPS features and the relative phase information is also combined with the Mel-Frequency Cepstral Coefficient (MFCC) and Modified Group Delay Coefficient(MGDC).To accomplish this project i am going to use “ASV spoof 2015 : Automatic Speaker Verification spoofing and countermeasures challenge” database.

# Introduction

In this study, we focus on whether speech is natural or spoofed.To detect spoofed speech from human speech, many features (e.g. magnitude spectrum,group delay, pitch,accent,modulation features) have been considered, pitch information, spectrum information was proposed to detect synthetic speech and cosine-normalized phase and MGDC phase spectrum features were used in voice converted speech detection.Modulation features were used in synthetic speech.In this project RPS features used to detect human speech or spoofed speech because the original phase information is entirely lost in spoofed speech.

# 2. Motivation

Recently many technology using speaker verification base security so, spoof detection is very important in these system,Motivation of this project is make system very effectiveness to words the spoofing detection and give better result with less error equality rate.

## 3. Application

RPS base feature extraction can be used in various kind of system as described below :

1. These system can be used in baking system.
2. Mobile,laptop and other device can be unlocked using this system
3. Telephones system can be verified using these system.
4. Speaker Verification

## 

## 4.System Architecture

In this project I implemented three features are RPS,Group Delay and Modified Group Delay Cepstral Coefficient and for classifier I used GMM and for evaluation I took dataset of **“ASV spoof 2017”** for DET curve I used NIST’s toolkit.

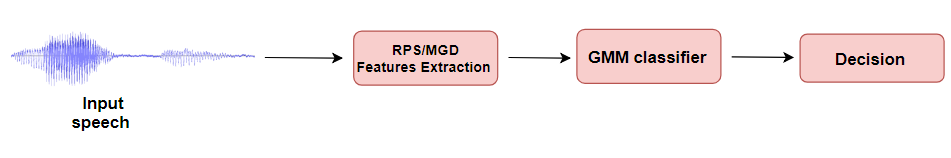
Here is the basic diagram for spoofing detection system :

Fig.4.1 Basic diagram of spoofing detection system

We give input as a speech which also known as test speech then extract various features like Mel Frequency Cepstral Coefficient, Relative Phase Shift,CFCC,MGDCC etc.In this project I have used Relative Phase Shift,Group Delay and Modified Group Delay Cepstral Coefficient,After extracting the features it goes to the GMM and In GMM already two model human and spoof already implemented and find the score.At the End we decide whether speech is natural or spoofed.

## 4.1 Overview of spoofing detection system using GMM

The flow chart of Gaussian Mixture Model is shown in figure 1.As per below diagram we have to give utterance of speech as input the extract features (can be MFCC, RPS or other) and then some selected features are given input in GMM.GMM is trained with natural speech and spoofed speech.so, test input feature will make decision.

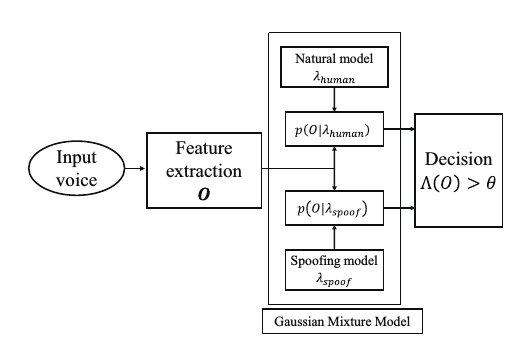
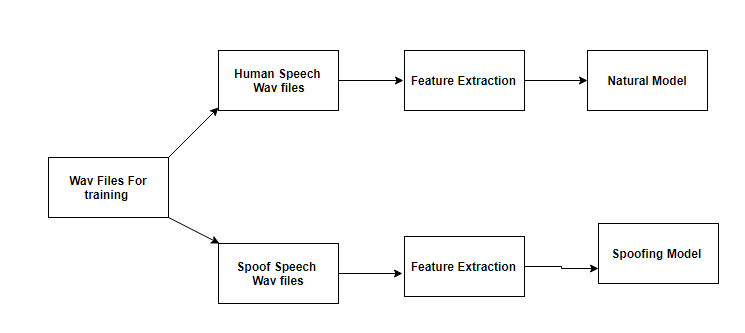


Figure 4.2. Flow chart of spoofing detection system

There are two phase in GMM:

**1.Training of GMM**

****

**Fig.4.3 Diagram of Training of GMM**

for train GMM I have used ASV spoof 2017 dataset in this dataset and protocol as per below

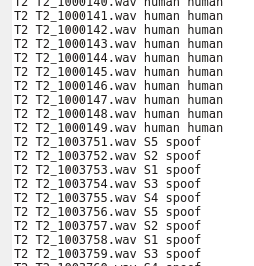


Fig.4.4 Training Protocol

Last two column are showing whether speech is natural or spoofed. Here s1 to s5 are known attack.As per diagram for all train wav file we will extract features and we created two different model for Human speech we got natural model and for spoof speech we got spoof model.

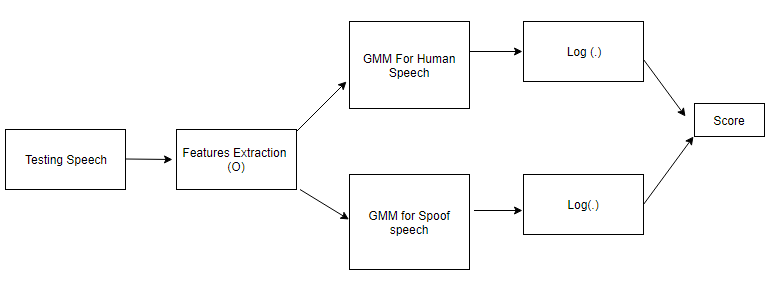
**2.Development (Testing)**

Fig. 4.5 Testing diagram

For Development of GMM, ASV spoof 2017 database snapshot as below :

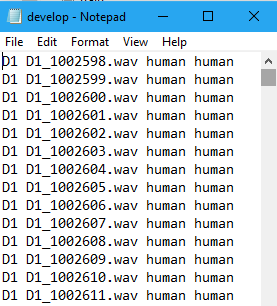


Fig. 4.6 Testing protocol

As above protocol, Test speech features are extracted and base on these features GMM will calculate GMMs for natural speech and and GMMs for spoof speech,After taking log we got log likelihood of GMMs.

The decision(score) about whether speech is natural or spoofed is based on log log likelihood ratio :

Λ(O) = log p(O|λhuman ) − log p(O|λspoof )..... eq(1)

Where O is feature vector of input speech, λhuman is GMM for natural and λspoof  GMM for spoofed speech.if GMM take multiple features as input then log likelihood ratio :

.... eq(2)

Where is the log likelihood ratio and denotes the weighting coefficients corresponding to features.

## 4.2 Relative Phase Shift

The definition of relative phase shift is The phase of certain base frequency kept constant and the phase of other frequency estimate relative to this base frequency.To understand consider a signal and it’s fourier transform and any fourier transform is written as terms of magnitude and phase as below

… eq(1)

The phase changes depending on the clipping position on the input speech even at the same frequency . To overcome this problem we kept certain base frequency as constant and calculate relative phase,For example, by setting the base frequency to 0. eq(3) is

.. eq(2)

So, other frequency fourier transform relative to base frequency can be written as below

… eq(3)

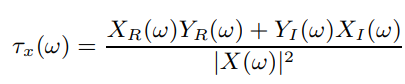
In eq(5) we will get normalized phase information can be written as below

… eq(4)

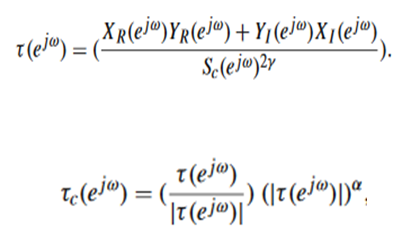
## 4.3 Group Delay

Group delay is define as a negative derivative of phase fourier transform with respect to frequency

Group delay can be directly derived from signal, consider a signal and Take fourier transform of both equation the we will get and this two quantity are complex.so group delay defined as below equation



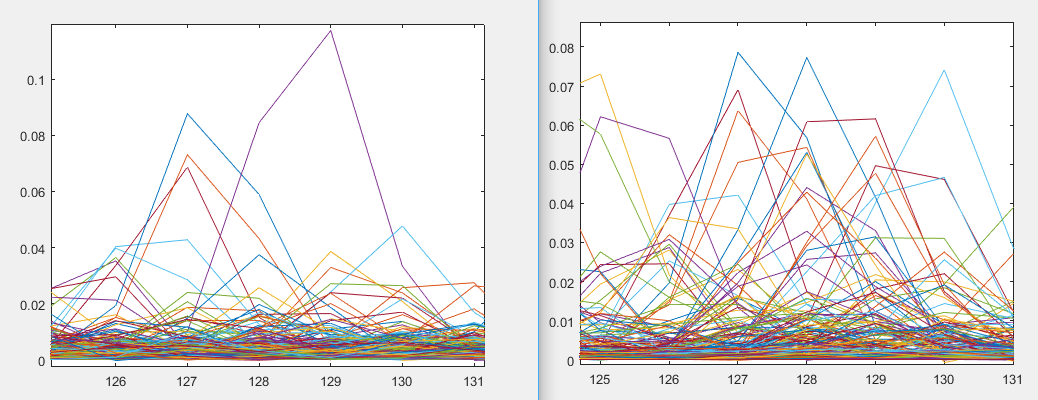
Here R is represent the real term and I is represent Imaginary term.Now we modified this equation we won’t touch numerator term we change numerator term as per below



Here and are control the shape of spectrum and these parameter make curve very smoothly.

## Novelty of My work

* Found alpha and gamma parameter for better result
* Comparison between group delay and modified group delay

****

Group Delay MGDCC

## 6. Algorithm Development

## 6.1 Algorithm Development of Relative phase shift

**6.1.1 Pitch tracking**

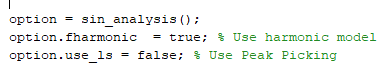
In RPS we first found fundamental frequency by using below function this algorithm is called a srh pitch tracking algorithm.The Summation of Residual Harmonics (SRH) method is a pitch tracker exploiting a spectral criterion on the harmonicity of the residual excitation signal.

the speech signal s(t) and the residual signal e(t) is obtained by inverse filtering. This whitening process has the advantage of removing the main contributions of both the noise and the vocal tract resonances. For each Hanning-windowed frame, covering several cycles of the resulting residual signal e(t), the amplitude spectrum E(f) is computed. E(f) has a relatively flat envelope and, for voiced segments of speech, presents peaks at the harmonics of the fundamental frequency F0. From this spectrum, and for each frequency in the range [F0,min, F0,max], the Summation of Residual Harmonics (SRH) is computed as:

For voicing speech male min frequency is 85 Hz and female max frequency is 255 Hz. so we will get fundamental frequency in these range.after getting fundamental frequency we can do sin analysis of signal.

**6.1.2 Sin Analysis**

In sin analysis estimate the parameter like amplitube,phase and frequency.For estimation we will use peak picking algorithm

￼

From previous step we got value of Fo .It will use in sin analysis as in below function

Frames contain phase so we can found RPS :

Hence, This way we can find rps features and we also found delta and double delta features.

## 6.2 Algorithm Development of Modified Group Delay cepstral coefficient

In Matlab we can implement function as we see in system architecture for modified group delay cepstrum:

Here alpha and gamma are parameter for shaping the curve.By trial and error i got alpha value as 0.2 and gamma value is 0.9 for good result.

## 6.3 Algorithm Development of Group Delay

It’s algorithm is same as modified group delay algorithm but take alpha and gamma value is equal to one.

## 7.pseudo code

### **7.1 pseudo code for RPS**

[RPS,option] = phase\_rps(frames, fs, option);

**Function phase\_rps algorithm:**

for n=1:size(f0s)

Ks = (0:size(frames(n).sins,2)-1)

pk = frames(n).sins(3,:); % 1.Amp , 2.fre, 3.The instantaneous phase

rps = pk - Ks.\*pk(1);

dctsignal = dct2(rps);

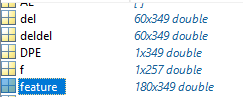
static = dctsignal(1:60,:); % First 60 features

del = deltas(static);

deldel = deltas(del);

feature = [static;del;deldel];

**Metrix Output :**

****

### 7.2 pseudo code for Modified Group Delay

**Modified Group delay cepstrum:**

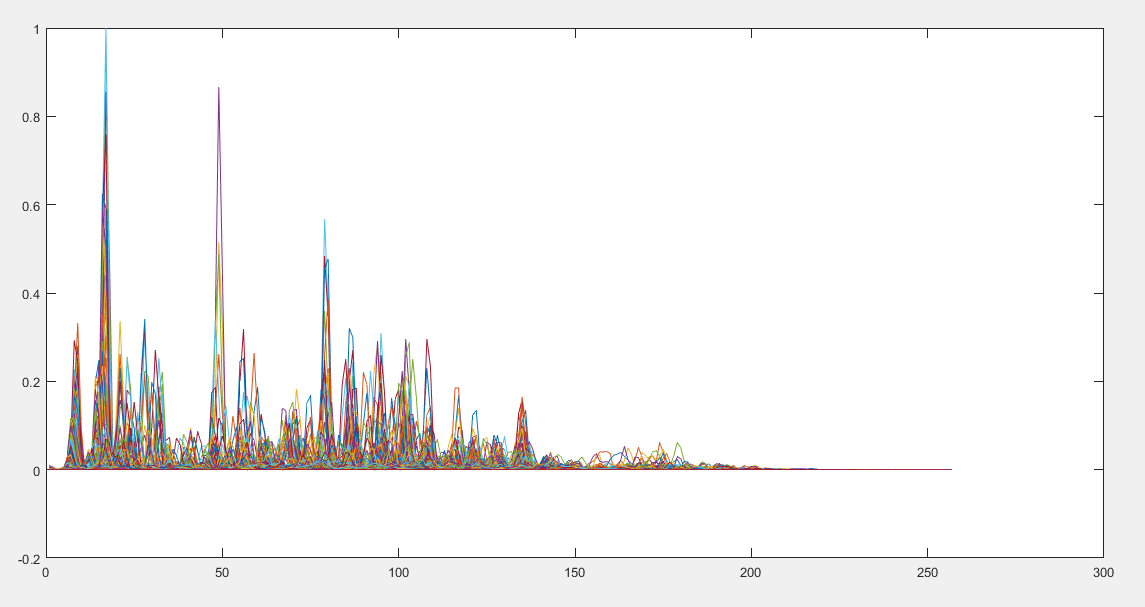
x(n)=speech signal

y(n)=nx(n)

grp\_phase1 = (real(x\_spec).\*real(y\_spec) + imag(y\_spec) .\* imag(x\_spec)) ./(exp(smooth\_spec).^ (2\*rho));

grp\_phase = (grp\_phase1 ./ abs(grp\_phase1)) .\* (abs(grp\_phase1).^ gamma);

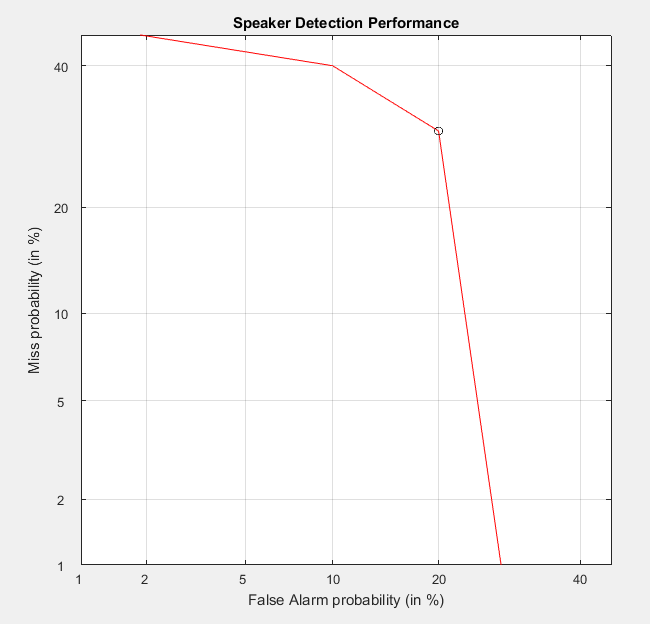
**Graph of modified Group delay:**

****

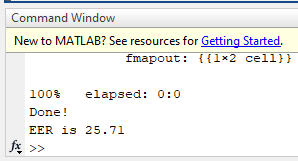
## 8. Result and Analysis

### **8.1 Result and Analysis for Relative Phase shift**

**DET :**

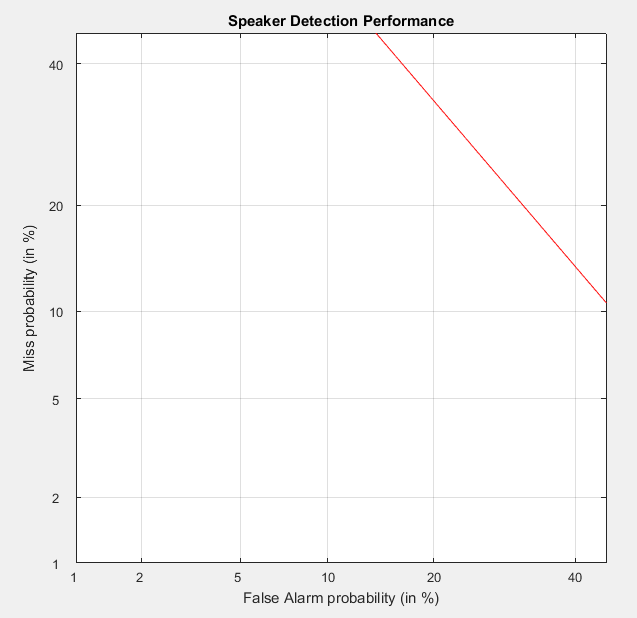
****

**EER:**

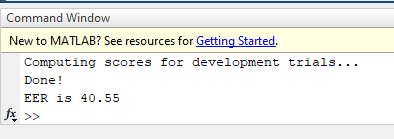
****

### 8.2 Result and Analysis for Group delay

**DET:**

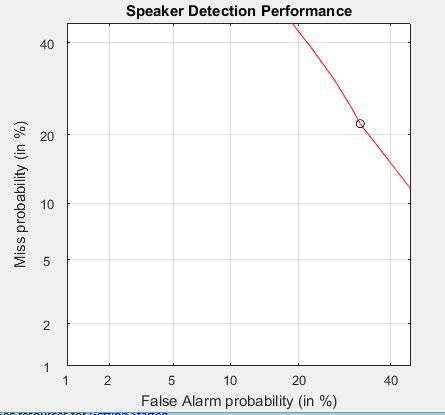
****

**EER:**

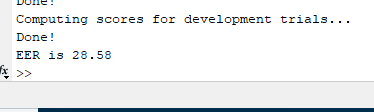
****

### 8.3 Result and Analysis for Modified Group Delay Cepstral Coefficient

**DET:**

****

**EER:**

****

### 8.4 comparison of result

|  |  |
| --- | --- |
| Feature Extraction Method | EER(Error Equality Rate) |
| Relative Phase Shift | 25.71 |
| Group Delay | 40.55 |
| Modified Group Delay Function | 28.58 |

Hence RPS is giving good result compare to modified group delay and group delay

## 9.Limitations

Limitation of this project is that we can’t extract RPS features for unvoiced speech

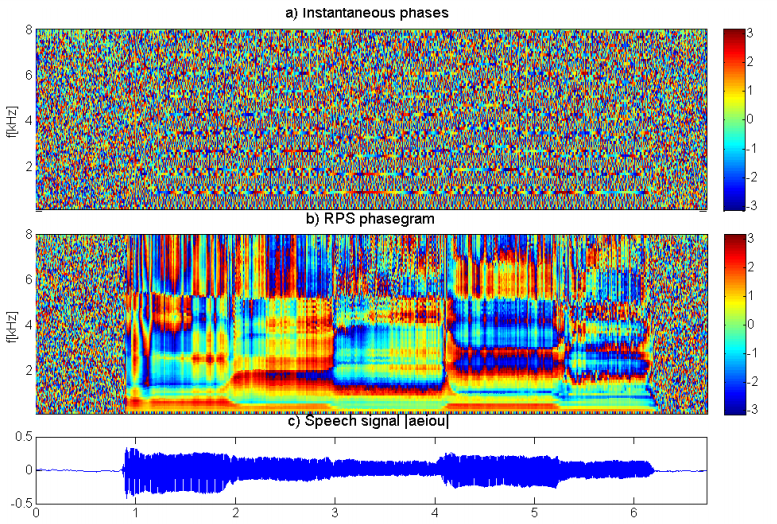


Fig a)instantaneous phases b)RPS phase gram c) speech signal

Above speech signal is utterance of vowels.This speech utterance contain voiced segment and unvoiced segment unvoiced means where zero crossing is very high.By analysing phase gram we can say that speech signal is not showing very good result so RPS features only for voiced segment.

## 

## 10.Future work

* Large Dataset would give very good results
* Relative phase shift and Modified Group Delay cepstral coefficient combination would give better result
* Combination of Relative Phase Shift and MFCC
* Combination of RPS + MFCC + MGDCC

## 

## 

## 11.References

* Longbiao Wang , Yohei Yoshida, Yuta Kawakami1and Seiichi Nakagawa, “Relative phase information for detecting human speech and spoofed speech”,ASVspoof Challenge 2015
* Longbiao Wang, Member, IEEE, Seiichi Nakagawa, Zhaofeng Zhang, Yohei Yoshida, and Yuta Kawakami, “Spoofing Speech Detection Using Modified Relative Phase Information”,IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, VOL. 11, NO. 4, JUNE 2017
* Hema a Murthy and B Yegnanarayana “Group delay functions and its applications in speech technology”,Sadhana Vol. 36, Part 5, October 2011, pp. 745–782. c Indian Academy of Sciences
* Simple representation of signal phase for harmonic speech models by I. Saratxaga, I. Herna´ez, D. Erro, E. Navas and J. Sa´nchez , ELECTRONICS LETTERS 26th March 2009 Vol. 45 No. 7
* Features and Classifiers for Replay Spoofing Attack Detection by Cemal Hanilci
* Speech Analysis/Synthesis Based on a Sinusoidal Representation by Robert J McAULY , THOMAS QUATIERI and IEEE Members,IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, VOL. ASSP-34, NO. 4, AUGUST 1986
* T.Drugman, A.Alwan, "Joint Robust Voicing Detection and Pitch Estimation based on residual harmonics” interspeech 2011
* Synthetic Speech Detection Using Phase Information by Ibon Saratxagaa , Jon Sanchez , Zhizheng Wub , Inma Hernaez
* ASVspoof 2017: Automatic Speaker Verification Spoofing and Countermeasures Challenge Evaluation Plan by Tomi Kinnunen , Nicholas Evans , Junichi Yamagishi , Kong Aik Lee , Md Sahidullah , Massimiliano Todisco , H´ector Delgado
* DETware\_v2.1 DET-Curve Plotting software for use with MATLAB https://www.nist.gov/file/65996