

# **Advanced Software Lab**

**It is an assignment on Latex**

Name: Deep Dey

Roll: 03

# Contents

<b>Abstract</b>	<b>5</b>
<b>1 Introduction</b>	<b>6</b>
<b>2 Procedures for Speaker Recognition</b>	<b>6</b>
<b>3 Speaker Recognition form Speech under Face Cover</b>	<b>8</b>
3.1 Related Works . . . . .	8
<b>4 Conclusion</b>	<b>10</b>
<b>5 Proposed Advancement</b>	<b>10</b>
<b>References</b>	<b>11</b>

## List of Figures

1	Beautiful Picture of Water Body . . . . .	6
2	Flowchart for speaker verification [2] . . . . .	7
3	Flowchart of Speaker Identification [2] . . . . .	7
4	[1] Long Term Average Spectrum of a male speaker under the influence of face covering elements . . . . .	8

## List of Tables

1	[1] Closed-Set Correct Speaker Identification rate in % . . . . .	9
2	[1] Confusion Matrix for Closed-Set masked identification in % . . . . .	9

## **Abstrunct**

Speech under face cover is a new aspect which is faced frequently by forensic speech experts. Face cover is used by people when they strive to hide their identity. Face covers not only blocks the scope of Face Recognition but also has huge effect on speaker recognition by the speech. Different types of face cover and different material absorbs speech energy differently resulting modulate speech of different kind. This leads to huge disadvantage in speaker recognition. Again there has been very little research work done, searching the effects in acoustics of speech through different face cover. Here is a survey on experiments done in Speaker Recognition under different Face Cover.

# 1 Introduction

Speech is foremost way of communication for human being. Speech not only is the way of communication but also very important to identify a person and get to know about their mental state.[1][2][3] Some experts can also find physical attributes of a person by his/her voice. So speech plays a vital role in human life and society.

As speech can be used as identifier of a person it is vastly used in different application as a Biometric security procedure. Since 1990's speaker recognition gained huge attraction and many research works have been performed. Now speaker recognition through Text Dependent (e.g. via password) or Text Independent (via acoustic qualities) system is more or less available. Forensic speech recognition uses automated system to identify the speaker along with expert listener.

But Speaker Recognition under the influence of face cover is a new aspect with increasing popularity since 2014.

## 2 Procedures for Speaker Recognition

Speaker recognition is broadly classified into two fields. Speaker Verification and Speaker Identification.

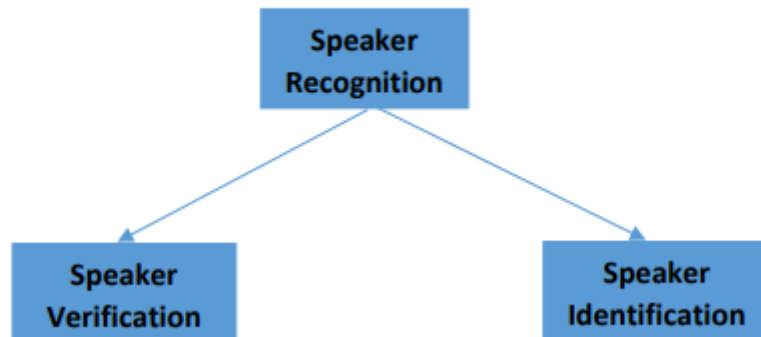


Figure 1: Beautiful Picture of Water Body

Speaker Verification: When a person claims to be someone known to system and the truth is verified. It is simple in Searching form database. It is just 1:1 mapping of user given data and database.

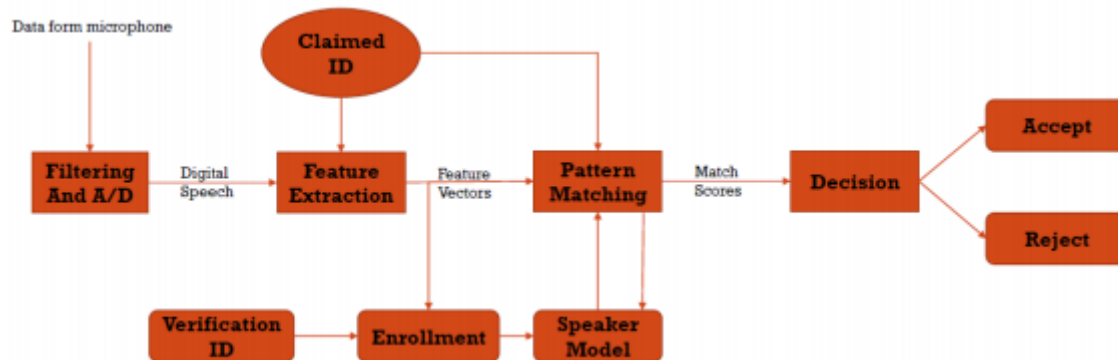


Figure 2: Flowchart for speaker verification [2]

Speaker Identification: When for some unknown person entire database is searched to find possible match. Here 1:N mapping is held between user data and database search.

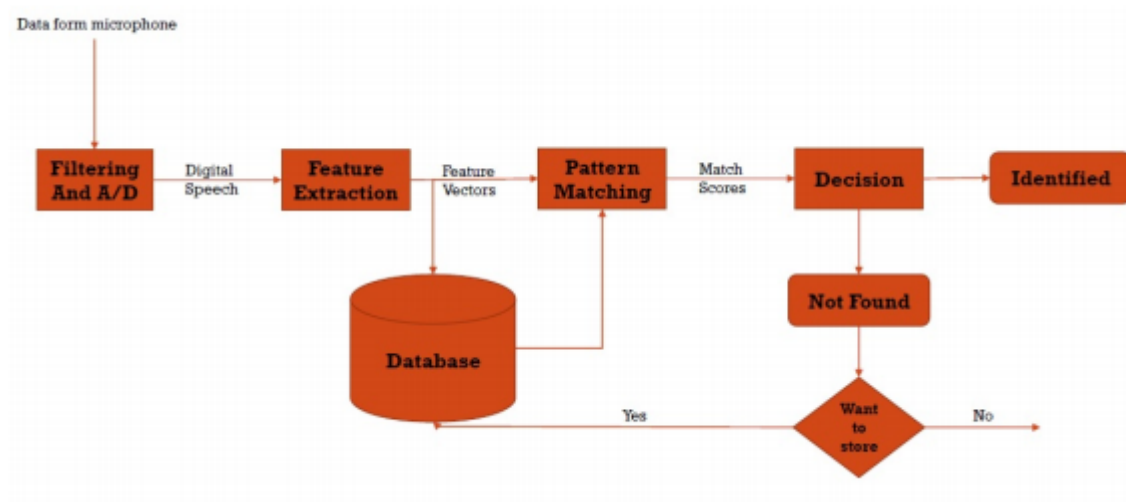


Figure 3: Flowchart of Speaker Identification [2]

### 3 Speaker Recognition form Speech under Face Cover

Human being is capable of identifying a person by his/her speech (i.e. voice quality). It is possible for us to recognize some person by their voice even though voice is molded via some materials which covers the face like Helmet, Mask and Scarf etc. But in the new area like voice recognition in Computer Science it is a very big challenge. Intentional voice modulation, in the field of forensic speech analysis, plays a vital role misleading automated voice recognition system or even expert listener. Voice modulation can be done by Imitation, speaking under face cover and synthesized speech.

Speech under face cover got the attention after the case of James Foley in Iraq [3]. The modification caused by face cover is generally of two types. *i*) Intrinsic (effected speech due to face cover) and *ii*) Extrinsic (signal absorption). The amount of face covered, degree of contact, restriction in jaw movement and type of the material are major factors in speech under face cover.

#### 3.1 Related Works

In June, 2015 Rahim Saeidi et al. has done a remarkable job in detecting the speaker under influence of face cover [1]. They studied the change in acoustic due to face cover with following result,

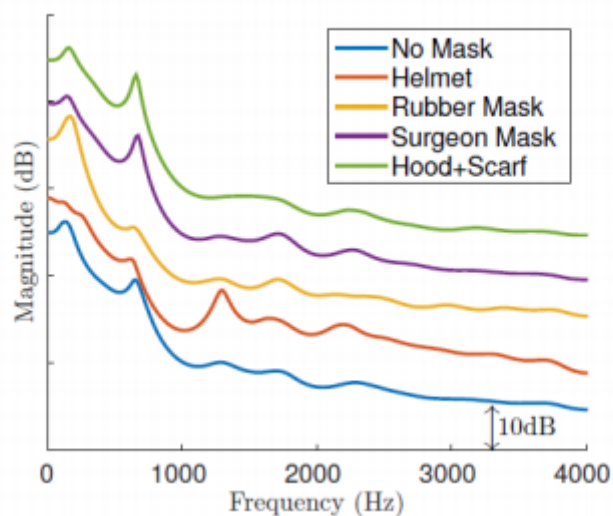


Figure 4: [1] Long Term Average Spectrum of a male speaker under the influence of face covering elements



They trained a gender dependent Universal Background Model [1] with 2048 samples. The results they got are summarized in the following table,

Face Cover	No Cover	Helmet	Rubber Mask	Surgeon Mask	Hood+Scarf
No Cover	<b>95.2</b>	94.9	88.6	94.2	93.3
Helmet	88.5	<b>97.7</b>	86.0	88.8	88.4
Rubber Mask	90.3	96.5	<b>97.1</b>	94.1	91.4
Surgeon Mask	95.1	96.7	90.1	<b>97.9</b>	95.6
Hood+Scarf	90.3	85.7	82.5	94.9	<b>97.0</b>
Number of Tests	793	574	543	626	568

Table 1: [1] Closed-Set Correct Speaker Identification rate in %

Face Cover	No Cover	Helmet	Rubber Mask	Surgeon Mask	Hood+Scarf
No Cover	73.9	28.7	16.9	40.7	24.5
Helmet	6.3	44.4	10.9	<b>9.9</b>	<b>4.6</b>
Rubber Mask	4.5	12.5	56.2	<b>7.2</b>	<b>5.8</b>
Surgeon Mask	4.7	<b>6.3</b>	<b>2.8</b>	17.1	10.4
Hood+Scarf	10.6	<b>8.0</b>	<b>13.3</b>	25.1	54.8

Table 2: [1] Confusion Matrix for Closed-Set masked identification in %

## 4 Conclusion

Speaker recognition is promising research area with lots of improvements to be done. It can be used as authentication key, again identification of a person is another issue.

Covering face is an event that frequently occurs in crime cases. For recent events like Iran [3] and similar incidents it is very important to discover speaker recognition from speech under face cover.

As it shows some limitations in the discussed works it will be noble to use different artificial intelligence base sophisticated techniques.

## 5 Proposed Advancement

Inspecting the results of the work by Rahim Saeidi et al. a remarkable success rate for identification is found. But this results shows some limitations

- (i) Matching in no cover scenario is inferior
- (ii) The training was done in a small variant of 8 speaker and 5 different frequencies
- (iii) The samples for matching were taken in suitable environment

So there is a lot of area left for improvement as

- (i) Improve matching for no-cover scenario;
- (ii) Improve matching of no-cover with different face cover;
- (iii) Train with large number of speaker;
- (iv) Use more face covers for different frequencies;
- (v) This may lead to accent and language identification with face cover.

## References

- [1] Mark Lutz. *Programming python.* ” O’Reilly Media, Inc.”, 2001.
- [2] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.
- [3] Guido Rossum. Python reference manual. 1995.