Replay Attack and Anti-Spoofing using F-ratio probing tool

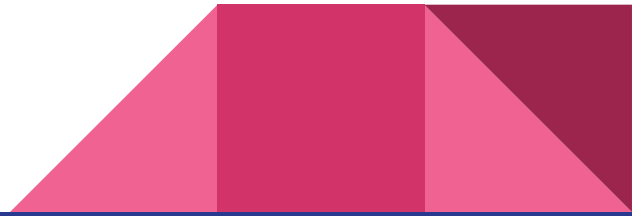
Name: Jay Shah
ID:201501071

Overview of Stage - II

- used the F-ratio metric as a probing tool to analyze the impact of various speech factors to replay detection.
- To discover which frequency bands are most discriminative for detecting replayed speech.
- Factors like Speaker, Speech Phrase, Recording Environment, Recording Device, Playback Device.
- Check generalizability between training data and development data

Algorithm

- Read wav file
- Windowing and FFT
- Applying filter bank
- Mean of frames on each filter bank
- Compute F-ratio on variability factor



Algorithm

$$F_i = \frac{(\mu_i^g - \mu_i^r)^2}{\frac{1}{N_g}\sum_{x_i \in C_g}(x_i - \mu_i^g)^2 + \frac{1}{N_r}\sum_{x_i \in C_r}(x_i - \mu_i^r)^2}$$

where x_i represents the value of the i-th Fbank of the speech frame x, and μ_i^g and μ_i^r are the means of x_i of all the frames of the genuine speech class and the replayed speech class, respectively. N_g and N_r are the number of frames of the two classes.

ASV Spoof 2017 Database

- Training Data

  Number of Genuine Utts. -> 1508

  Number of Spoofed Utts. -> 1508

- Development Data

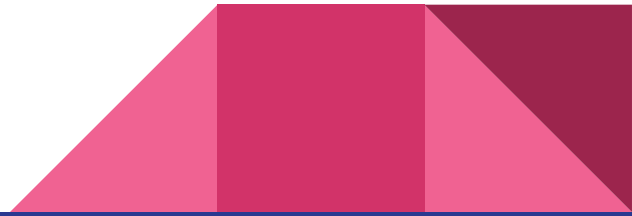
  Number of Genuine Utts. -> 760

  Number of Spoofed Utts. -> 950

ASV Spoof 2017 Database

File contains:-

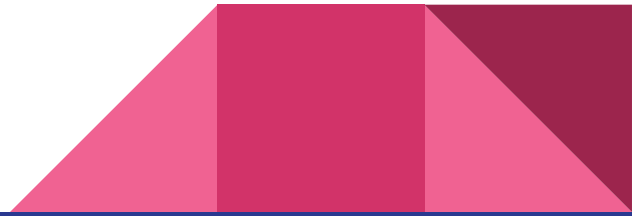
- Wave file
- Speech type (Genuine, Spoof)
- Speaker(Variability Factor)
- Phrase(Variability Factor)
- Environment(Variability Factor)
- Playback Device(Variability Factor)
- Recording Device(Variability Factor)



ASV Spoof 2017 Database

Training:-

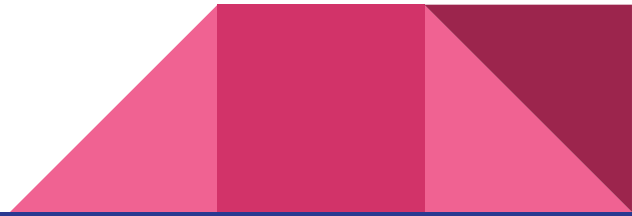
Number of Speakers	->	10
Number of Phrase	->	10
Number of Different Environment	->	2
Number of Playback Devices	->	3
Number of Recording Devices	->	1

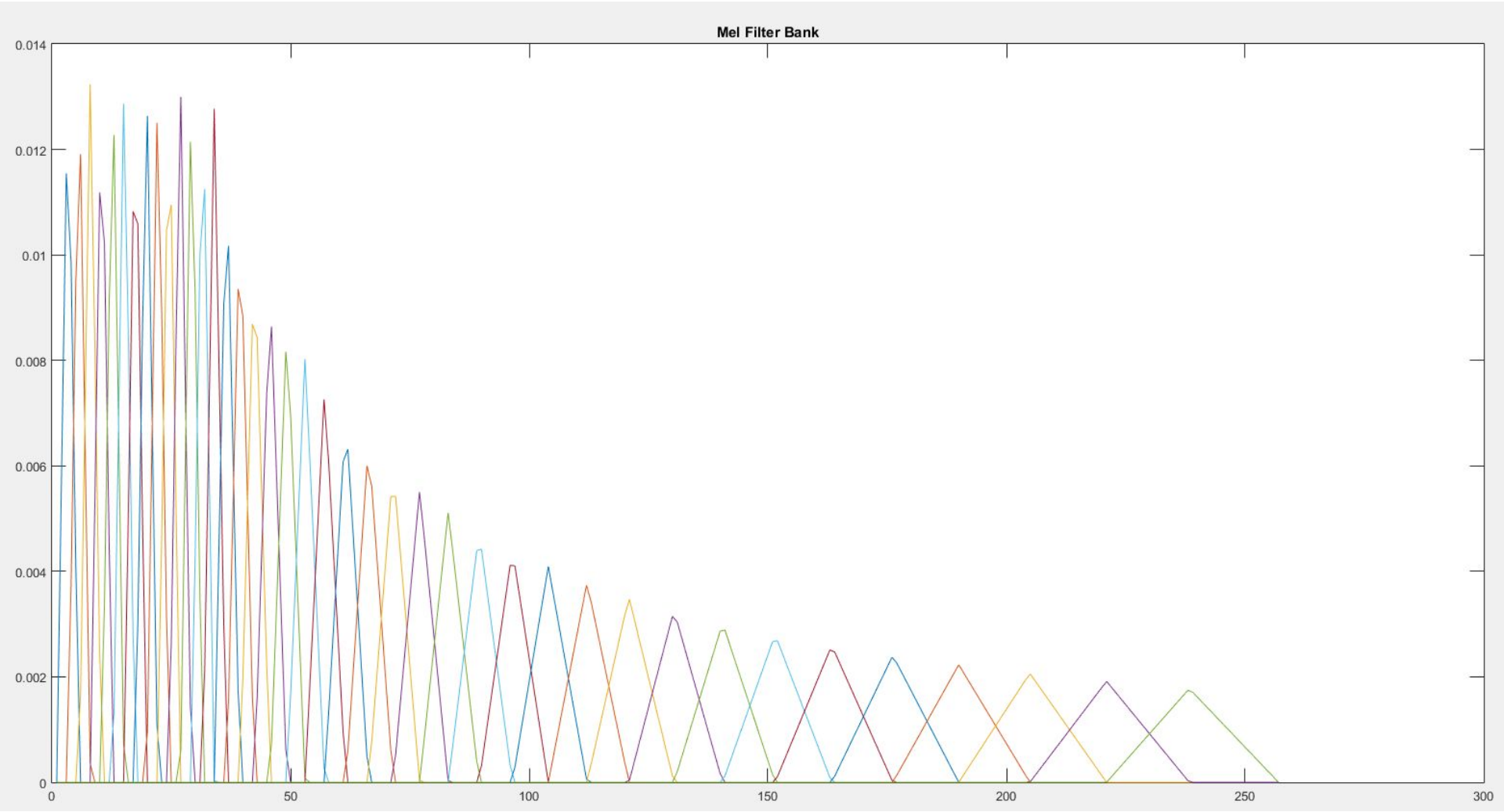


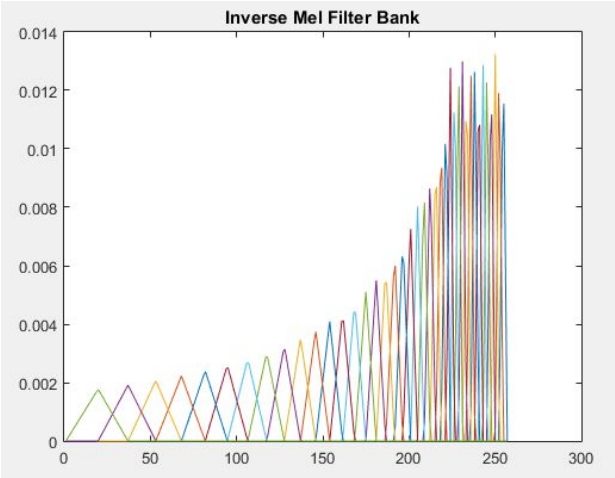
ASV Spoof 2017 Database

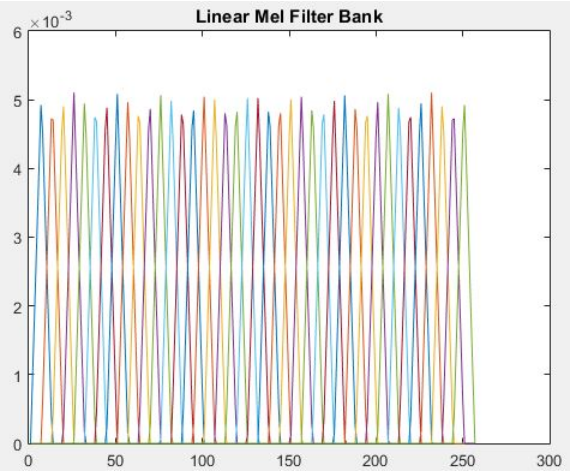
Development:-

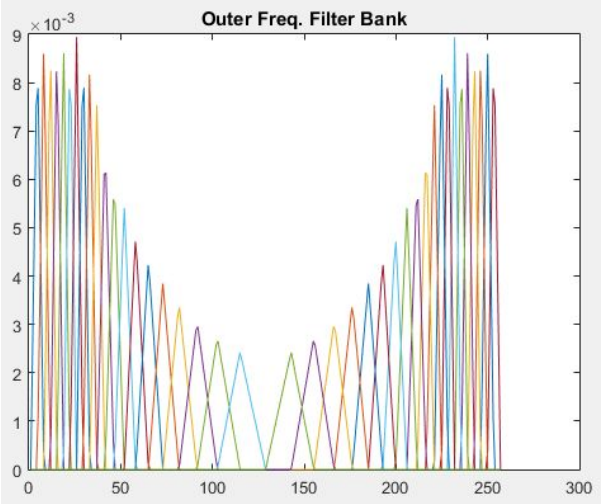
Number of Speakers	->	8
Number of Phrase	->	10
Number of Different Environment	->	6
Number of Playback Devices	->	6
Number of Recording Devices	->	7

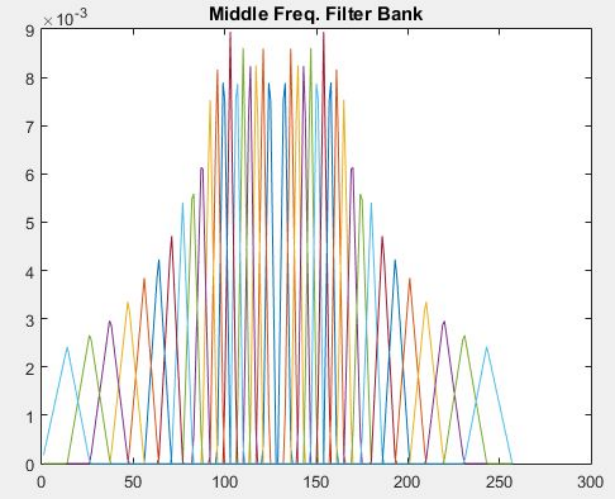



Mel Filter Bank
0.014
0.012
0.01
0.008
0.006
0.004
0.002
0
0
50
100
150
200
250
300

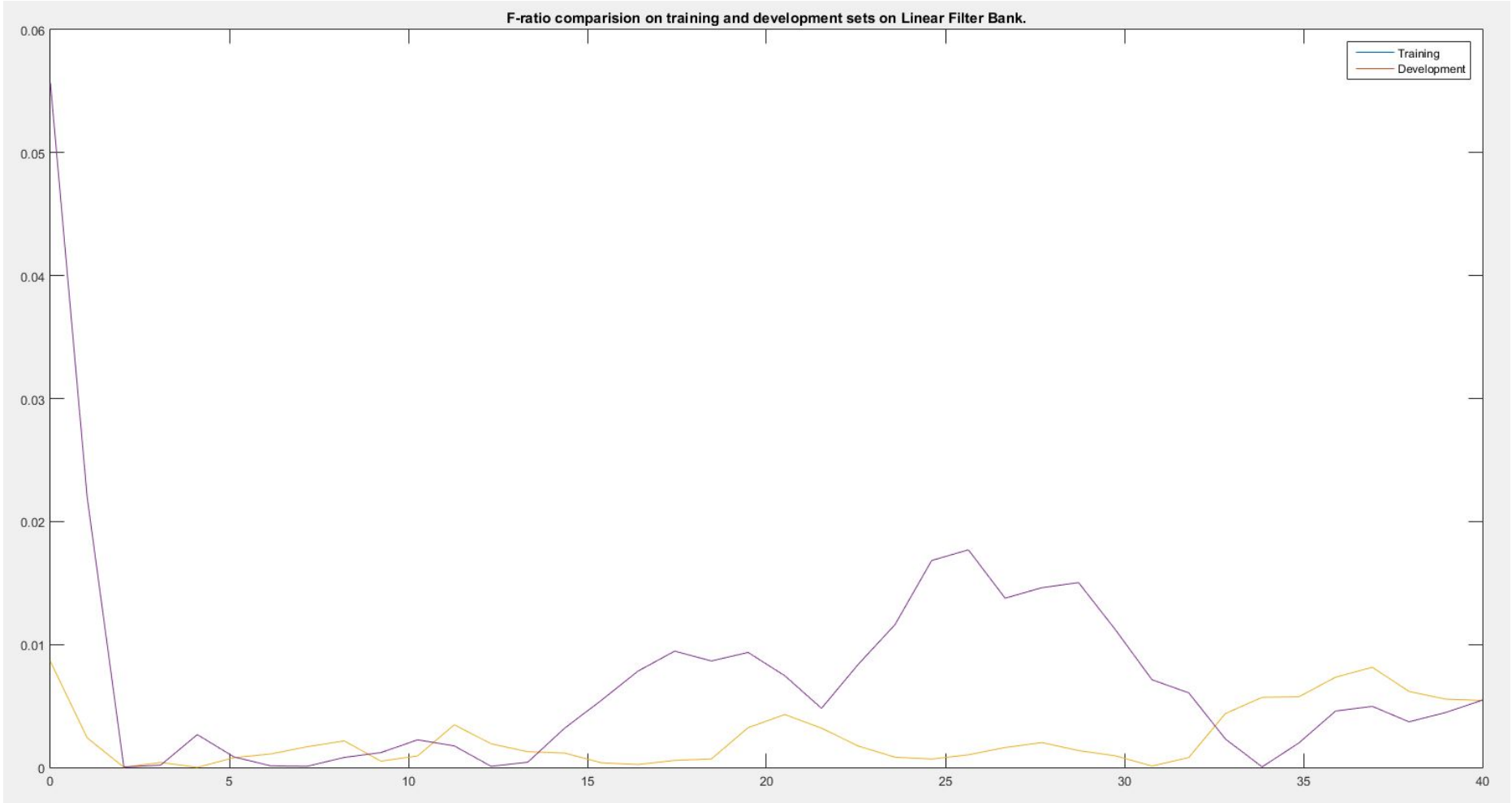

Inverse Mel Filter Bank
0.014
0.012
0.01
0.008
0.006
0.004
0.002
0
0
50
100
150
200
250
300

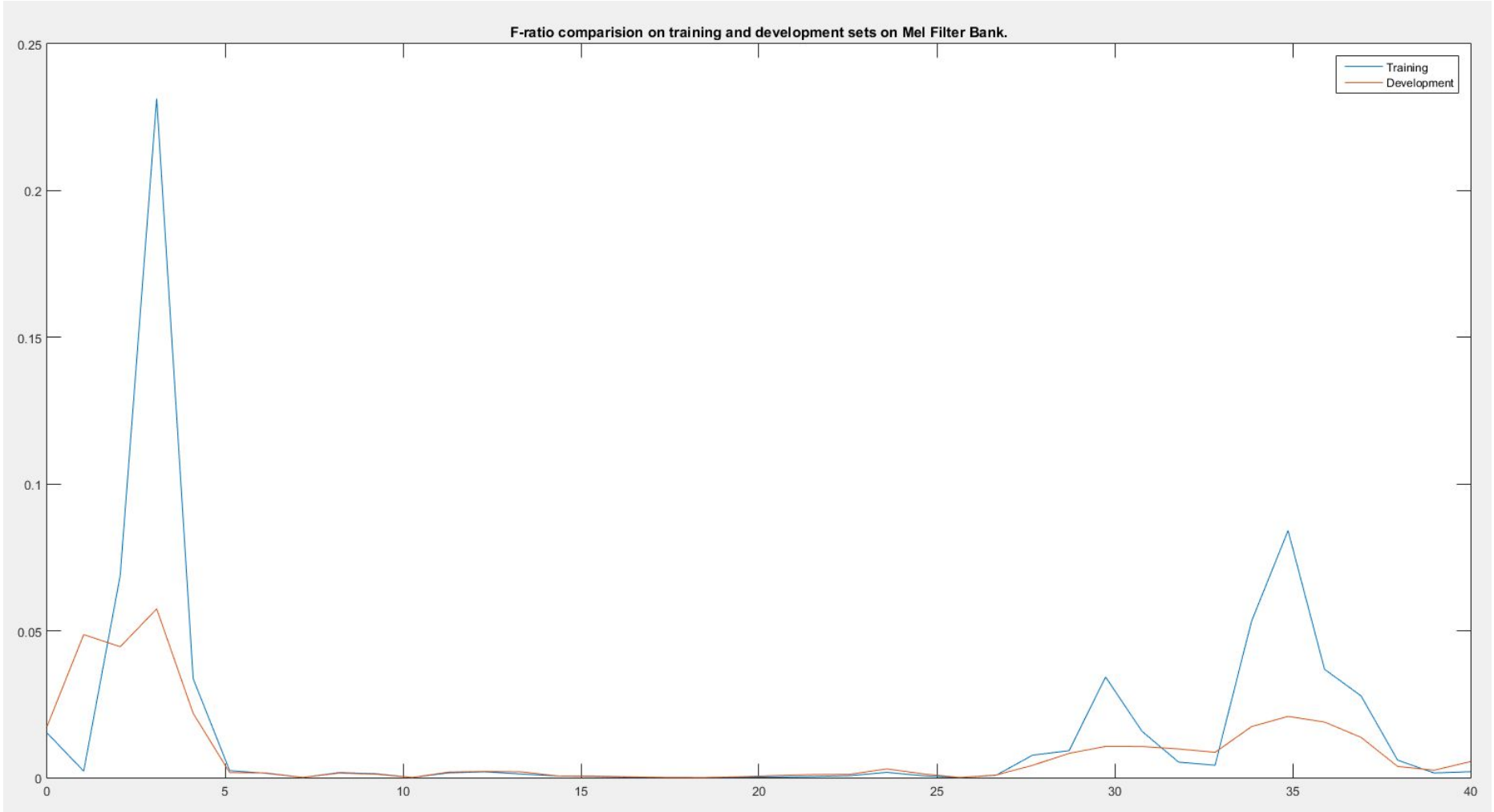

Linear Mel Filter Bank
×10-3
6
5
4
3
2
1
0
0
50
100
150
200
250
300

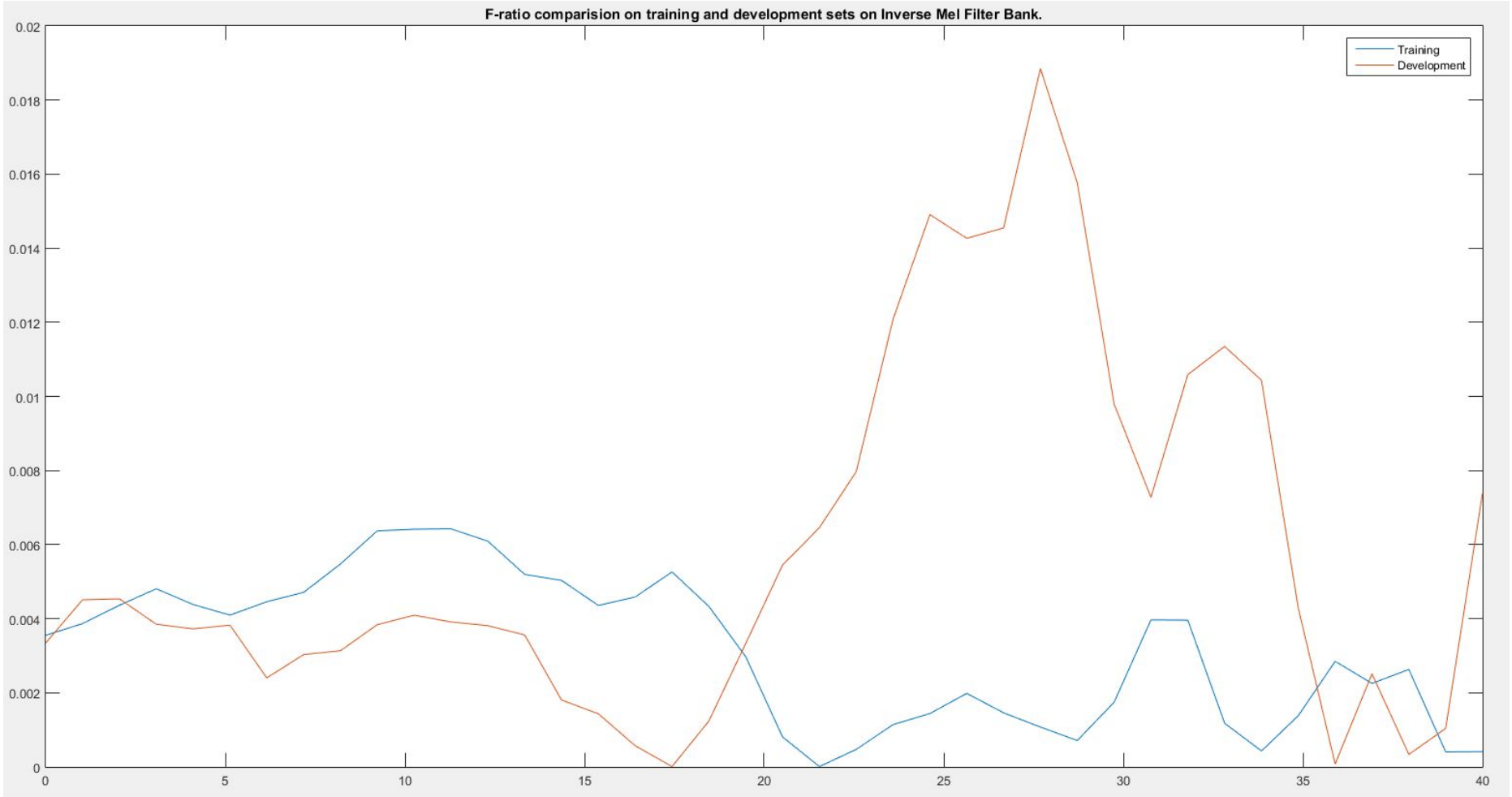

Outer Freq. Filter Bank
×10-3
9
8
7
6
5
4
3
2
1
0
0
50
100
150
200
250
300

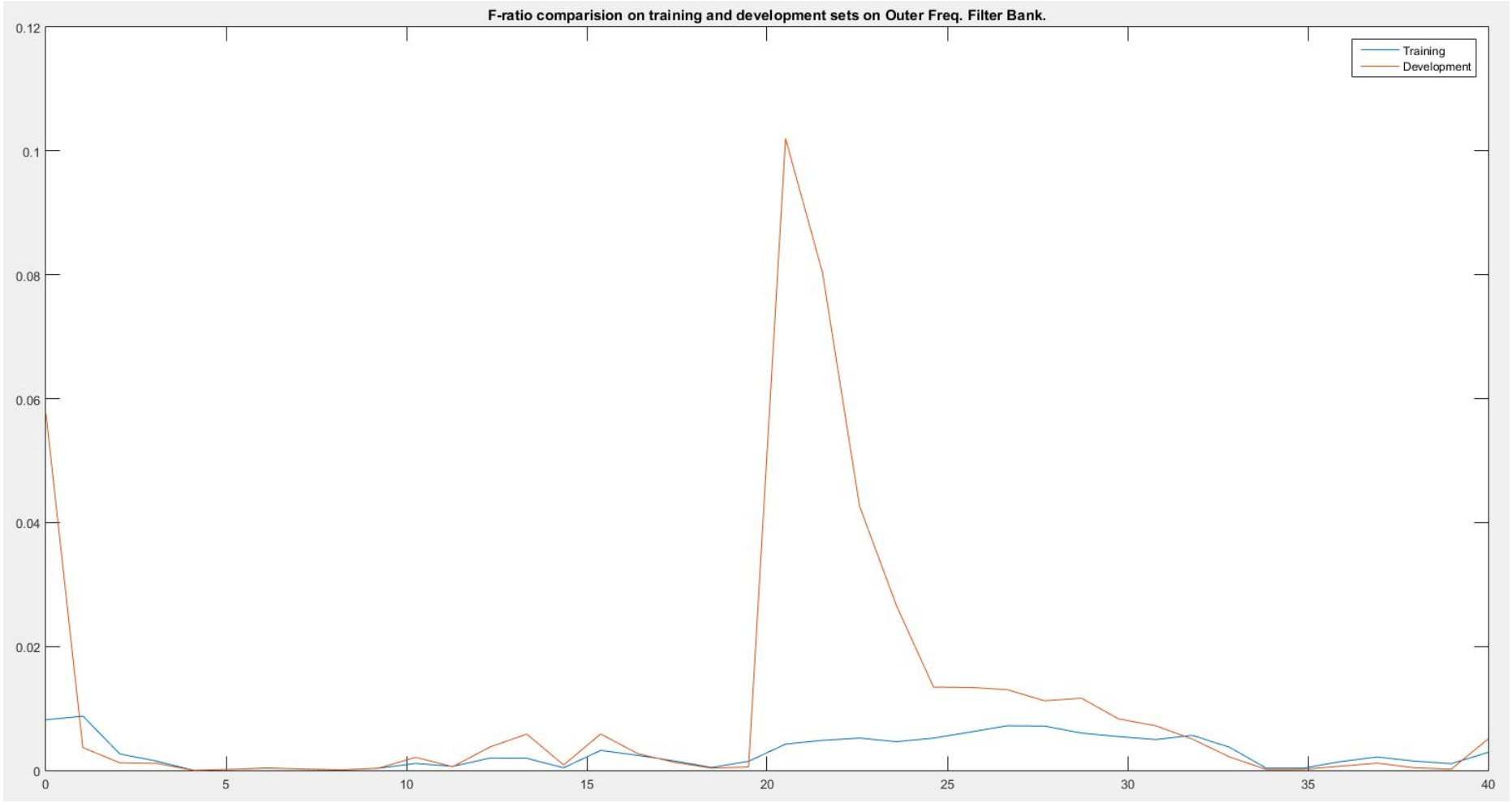

Middle Freq. Filter Bank
×10-3
9
8
7
6
5
4
3
2
1
0
0
50
100
150
200
250
300

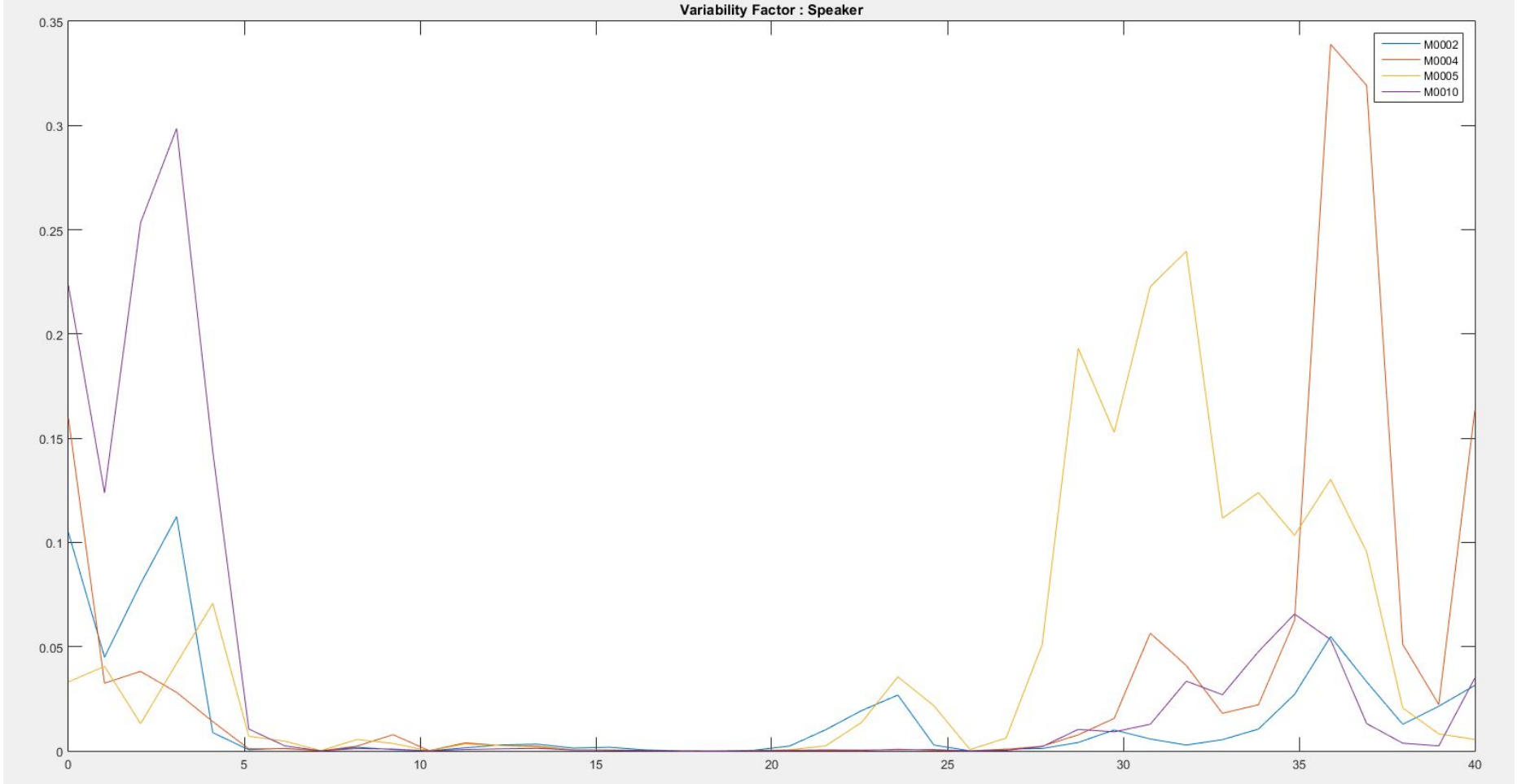
Results

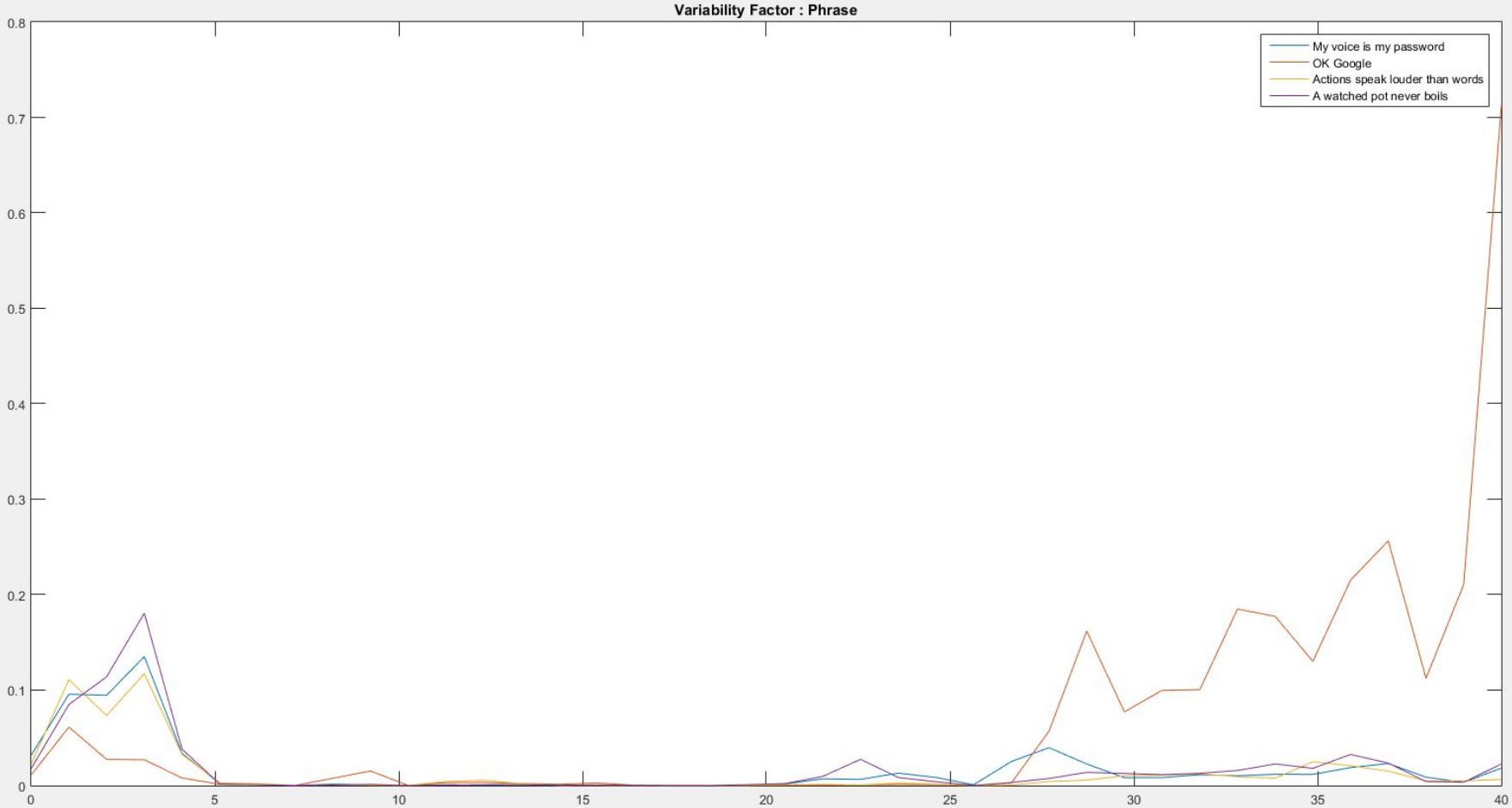

F-ratio comparision on training and development sets on Linear Filter Bank.
Training
Development
0.06
0.05
0.04
0.03
0.02
0.01
0
0
5
10
15
20
25
30
35
40

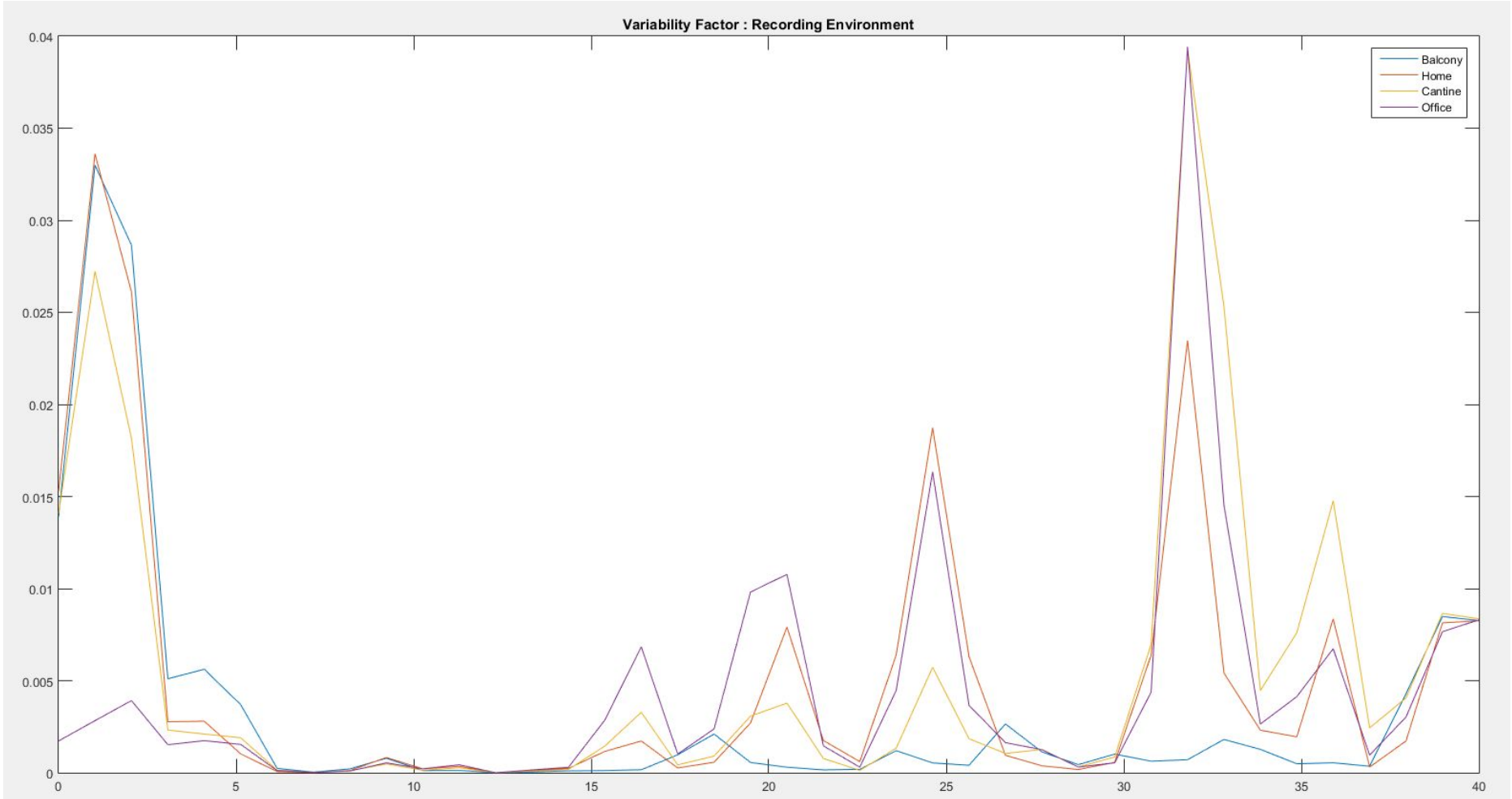

F-ratio comparision on training and development sets on Mel Filter Bank.
Training
Development
0.25
0.2
0.15
0.1
0.05
0
0
5
10
15
20
25
30
35
40

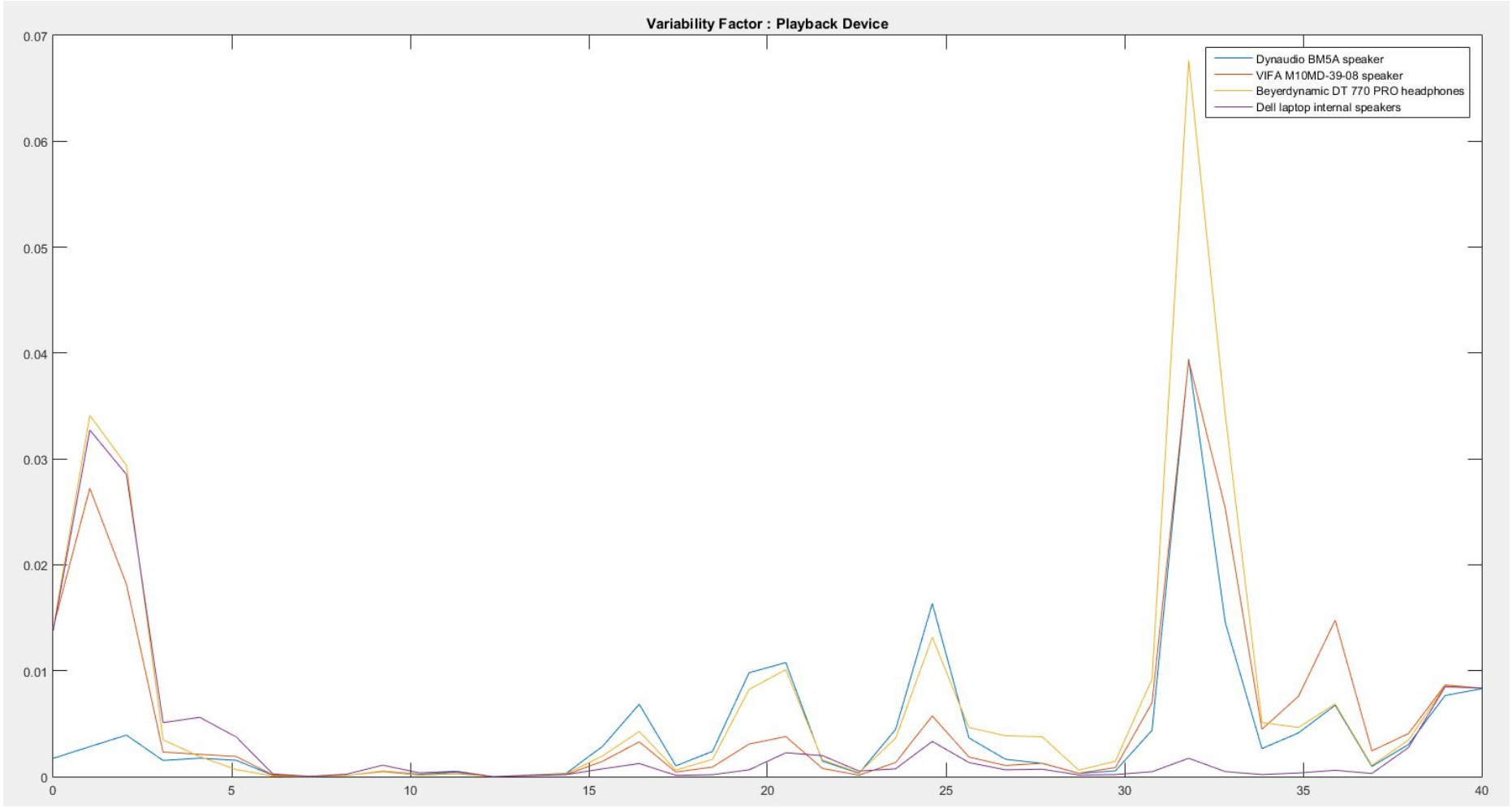

F-ratio comparision on training and development sets on Inverse Mel Filter Bank.
Training
Development
0.02
0.018
0.016
0.014
0.012
0.01
0.008
0.006
0.004
0.002
0
0
5
10
15
20
25
30
35
40

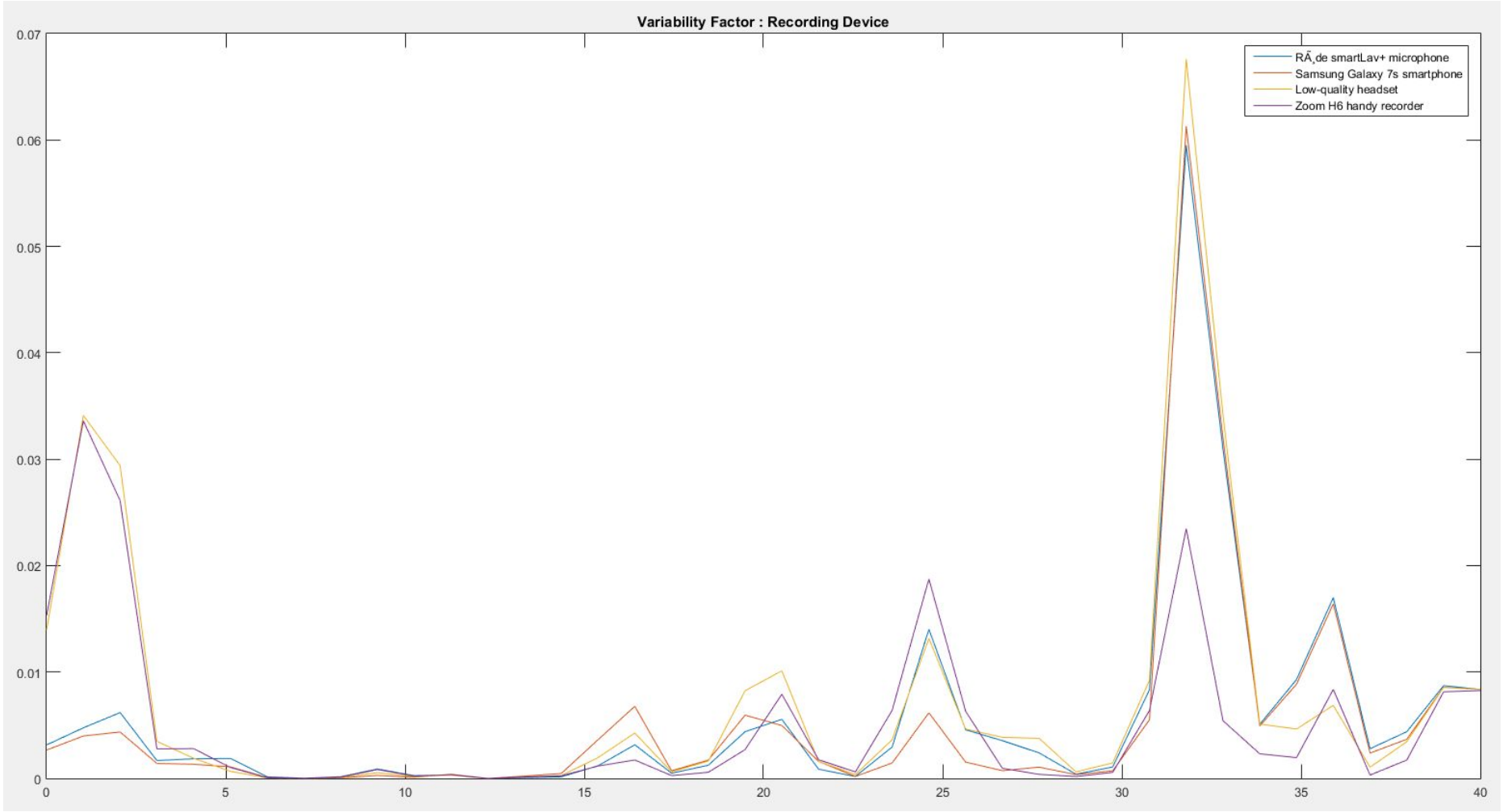

F-ratio comparision on training and development sets on Outer Freq. Filter Bank.
Training
Development
0.12
0.1
0.08
0.06
0.04
0.02
0
0
5
10
15
20
25
30
35
40


Variability Factor : Speaker
M0002
M0004
M0005
M0010
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
5
10
15
20
25
30
35
40


Variability Factor : Phrase
My voice is my password
OK Google
Actions speak louder than words
A watched pot never boils
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
5
10
15
20
25
30
35
40

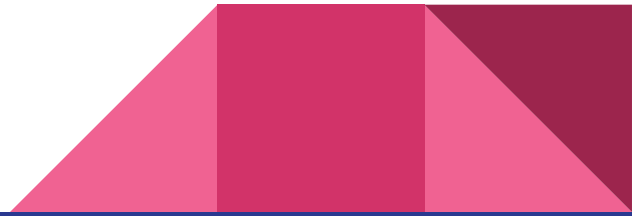

Variability Factor : Recording Environment
Balcony
Home
Cantine
Office
0.04
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
0
5
10
15
20
25
30
35
40

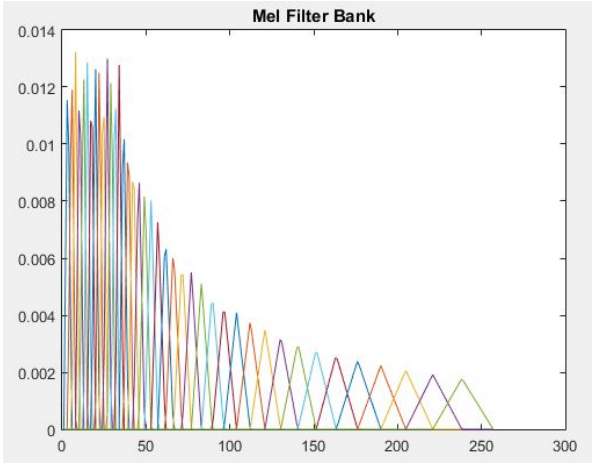

Variability Factor : Playback Device
Dynaudio BM5A speaker
VIFA M10MD-39-08 speaker
Beyerdynamic DT 770 PRO headphones
Dell laptop internal speakers
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
5
10
15
20
25
30
35
40


Variability Factor : Recording Device
RÃ¸de smartLav+ microphone
Samsung Galaxy 7s smartphone
Low-quality headset
Zoom H6 handy recorder
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
5
10
15
20
25
30
35
40

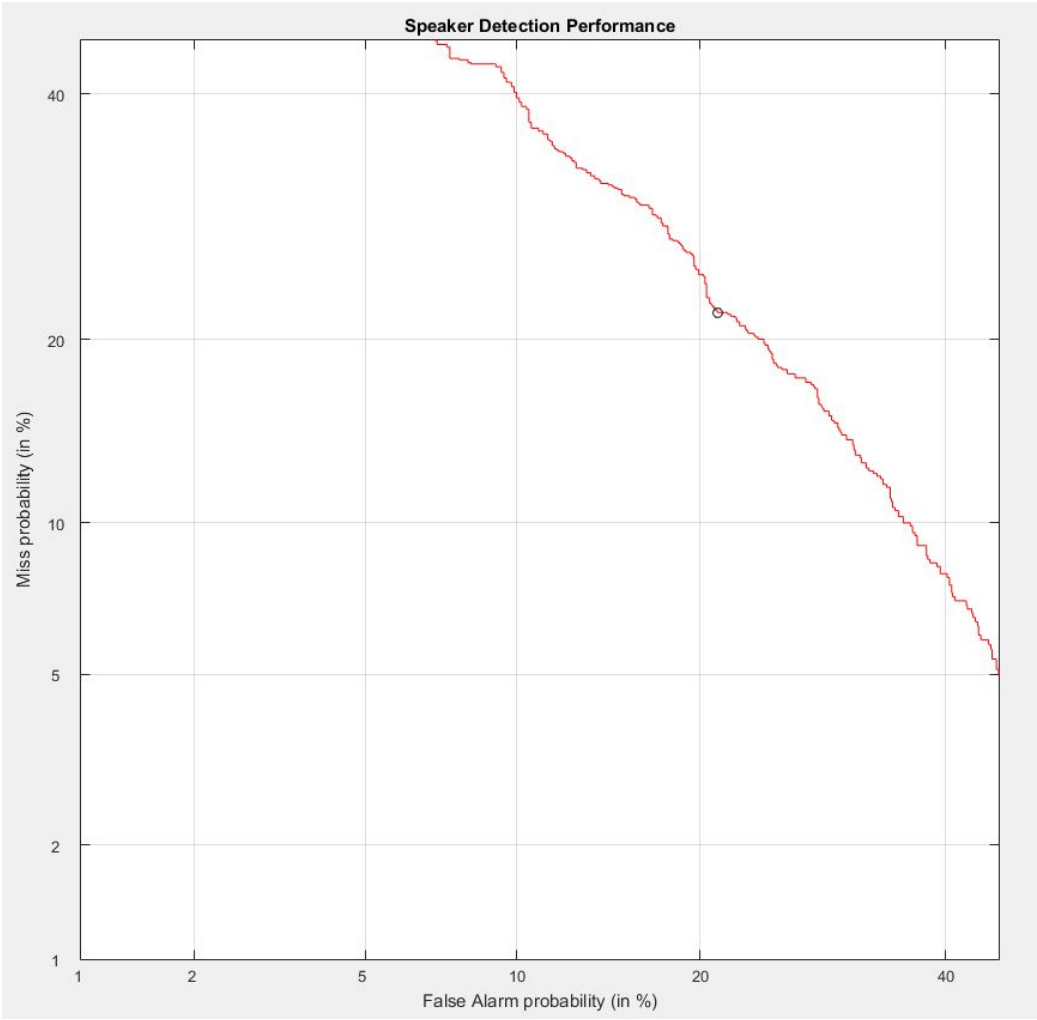
GMM Classifier

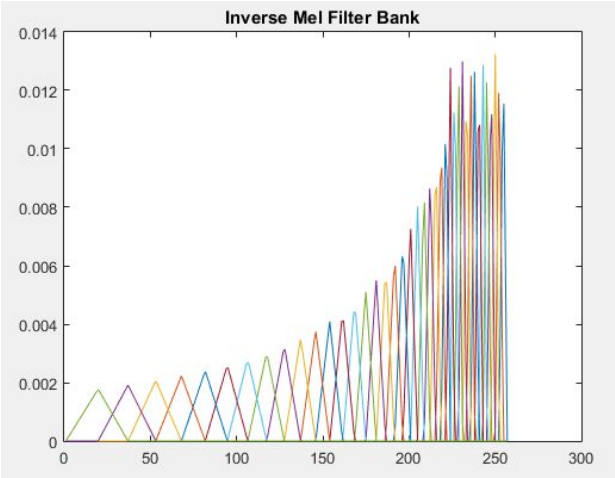
- Feature Extraction
- Train GMM on Genuine and Spoof Speech on Training Data Set
- For Development Data Set for every file take llk(log likelihood ratio)
- Score = llk_Genuine - llk_Spoof
- Human Score, Spoof Score (rocch)
- P_miss, P_falseAlaram
- DET
- EER



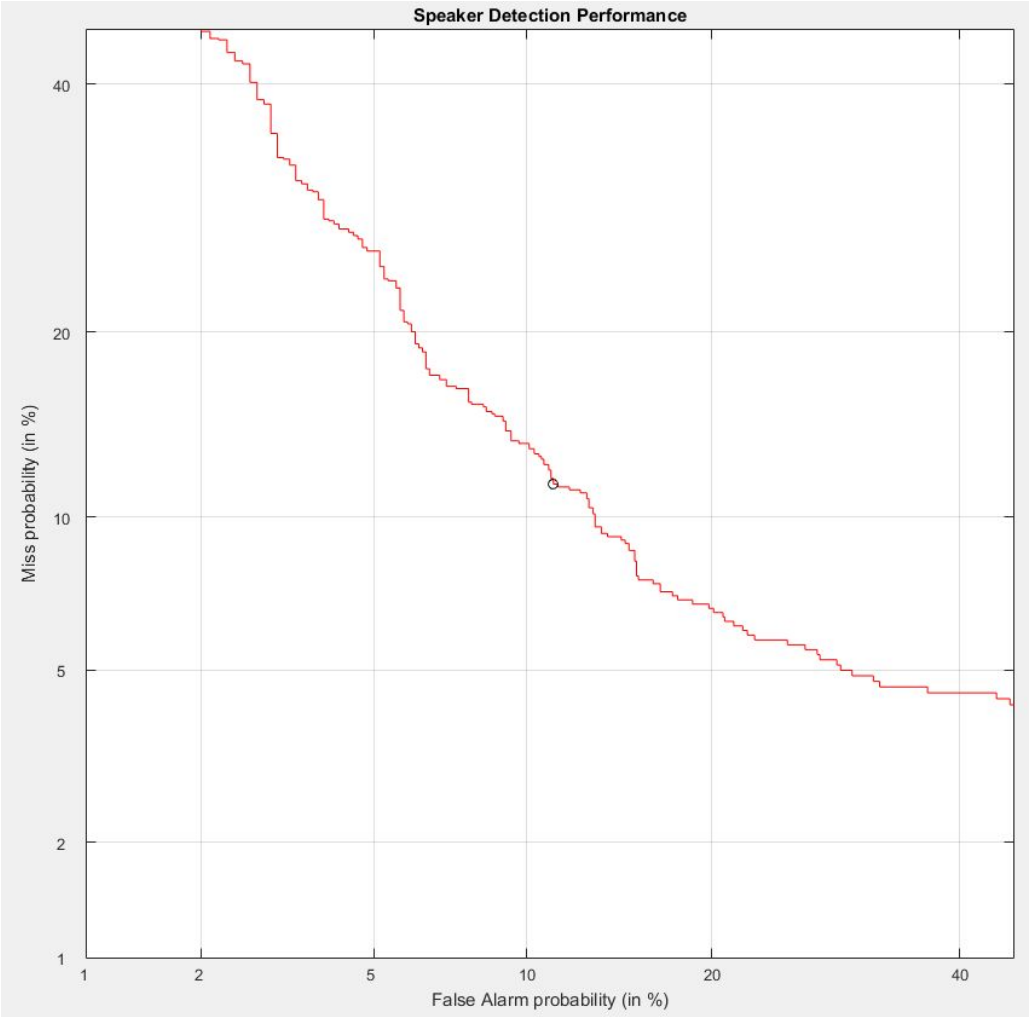


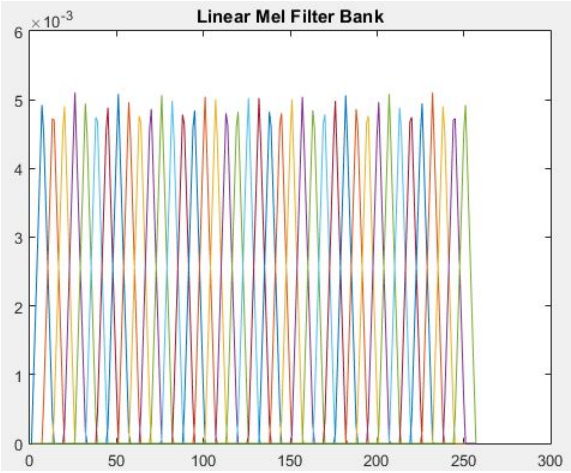

```
Computing scores for development trials...
EER is 21.53
```



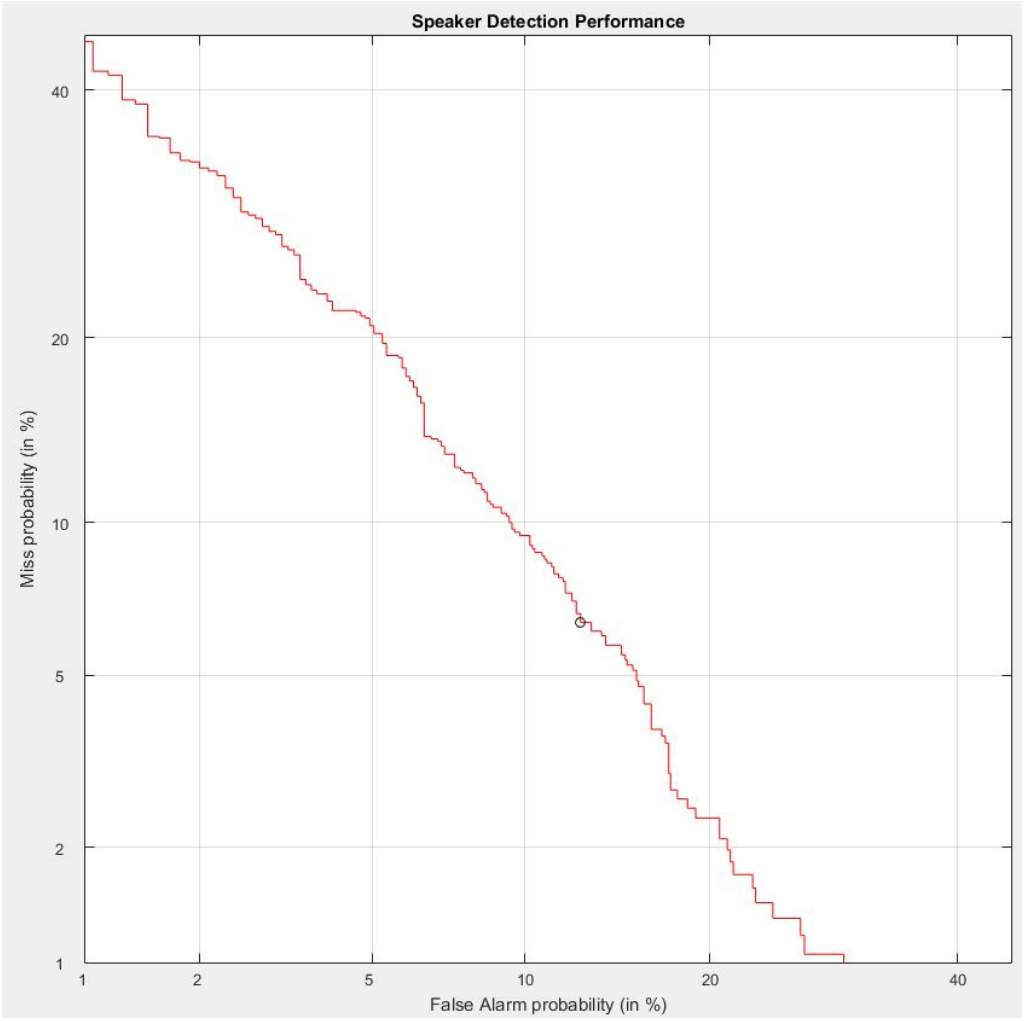


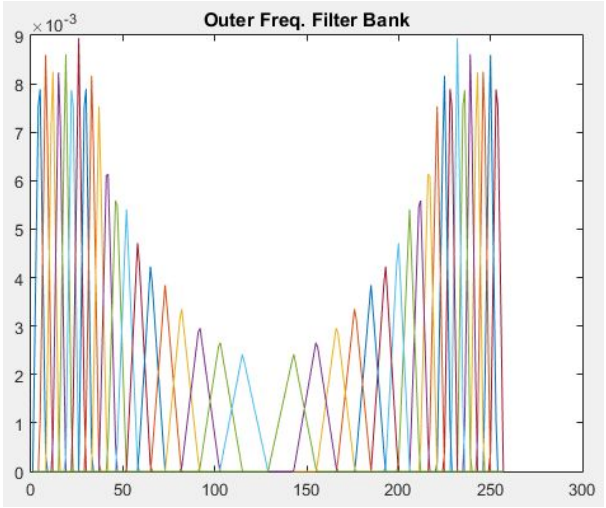

```
Computing scores for development trials...
EER is 11.31
```



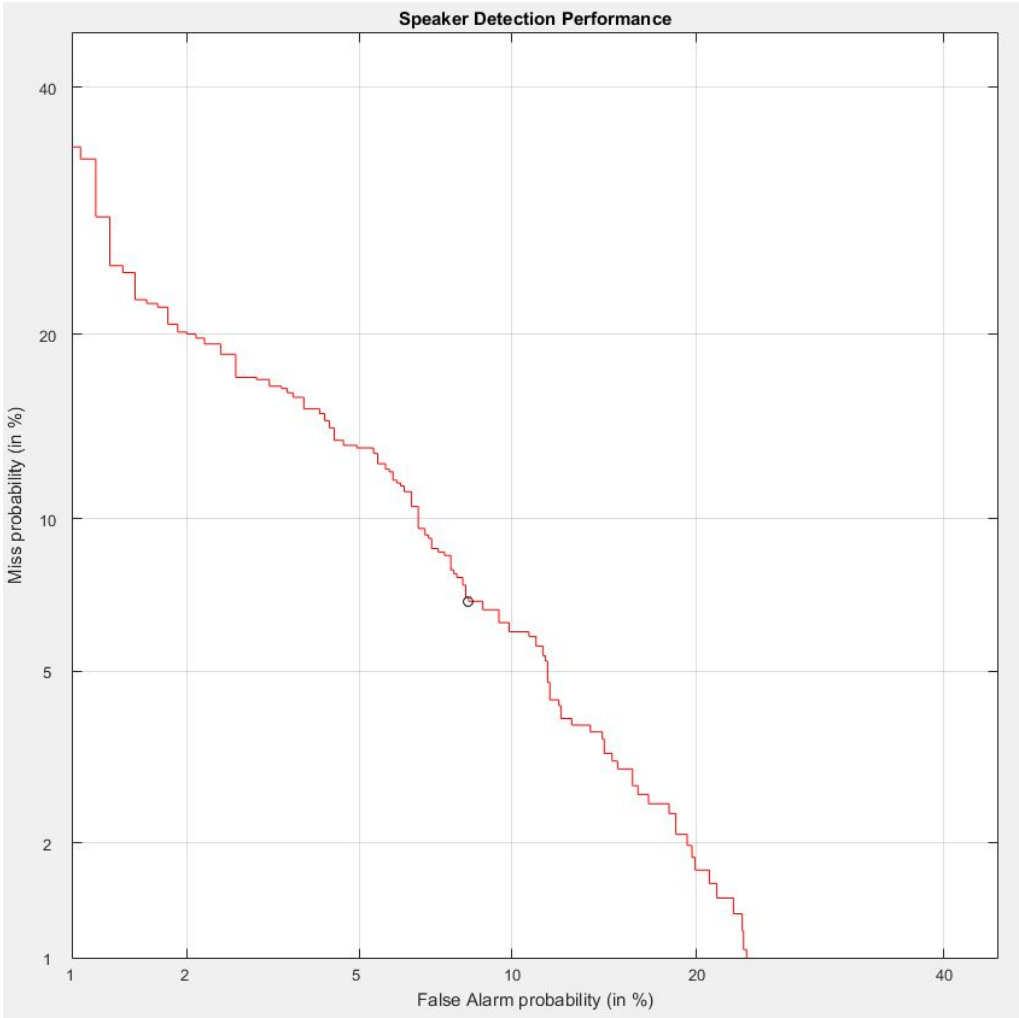


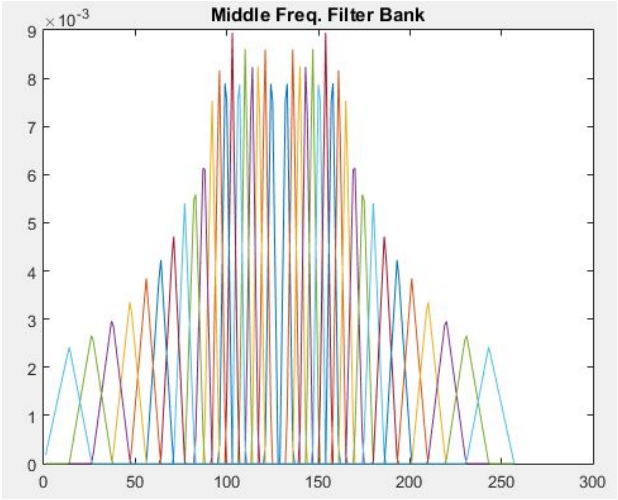

```
Computing scores for development trials...
EER is 9.59
```



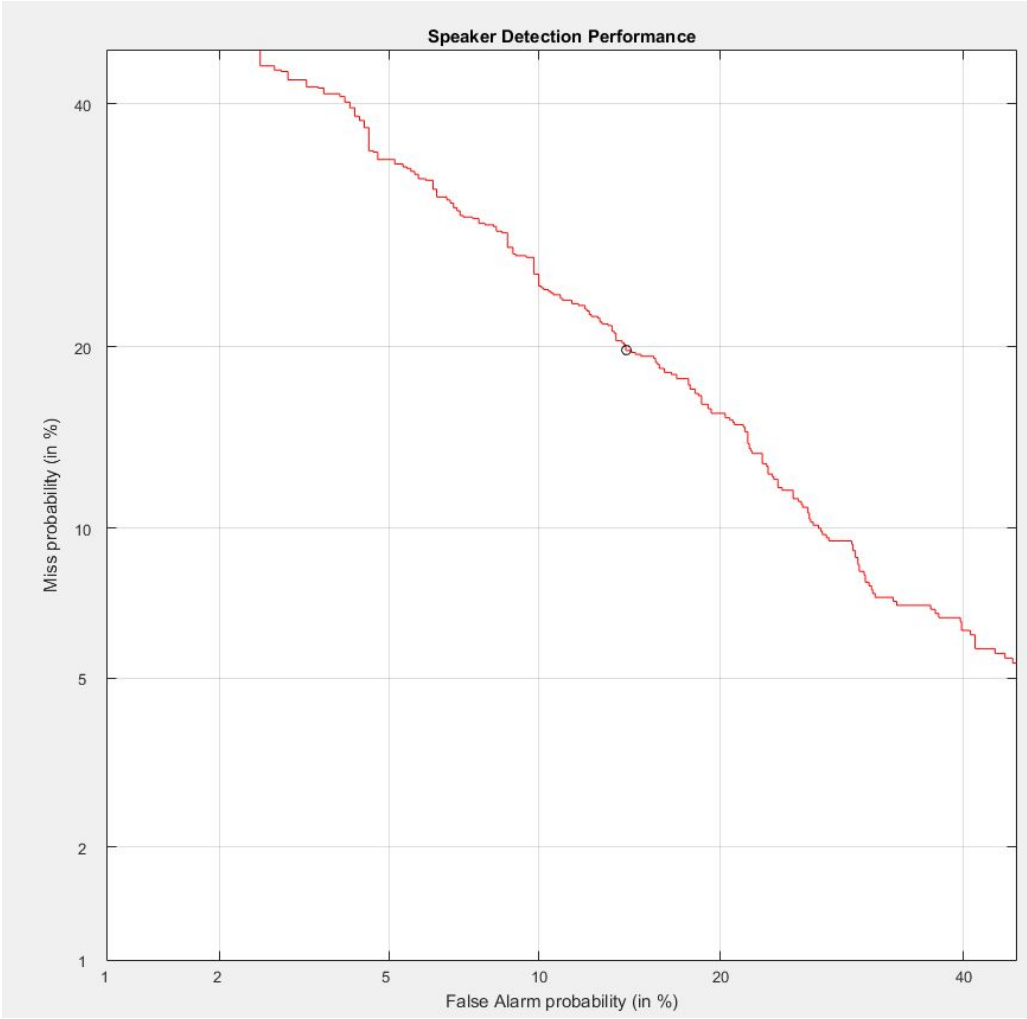



```
Computing scores for development trials...
EER is 7.76
```



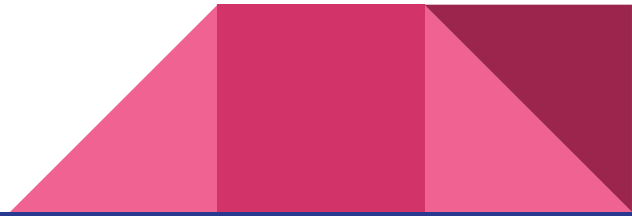



```
Computing scores for development trials...
EER is 17.28
```



Limitation of work

- Can not find discriminative information for replay detection for which frequency bands.
- More variation of F-ratio higher risk of overfitting.



References

- Lantian Li, Yixiang Chen, Dong Wang, Thomas Fang Zheng, ” A Study on Replay Attack and Anti-Spoofing for Automatic Speaker Verification.”INTERSPEECH 2017
- ASVSpoof 2017 dataset http://www.asvspoof.org/
- ASVSpoof 2017 baseline http://www.asvspoof.org/
- DETware_v2.1.tar.gz DET-Curve Plotting software for use with MATLAB

Thank You