Speaker Verification Anti-Spoofing : Replay and Imitation Attacks

Bhusan Chettri,
Supervised by: Dr. Bob L. Sturm and Dr. Ioannis Patras

Machine Listening Lab,
Queen Mary University of London

14 June, 2017

Outline

Automatic speaker recognition

Spoofing challenge

Experiments

Research goals and plans

Automatic speaker verification (ASV) and identification

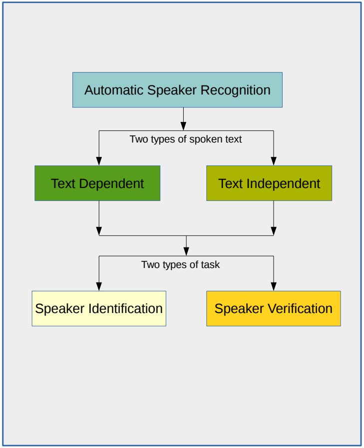


Figure 1 – Overview of Automatic Speaker Recognition Systems

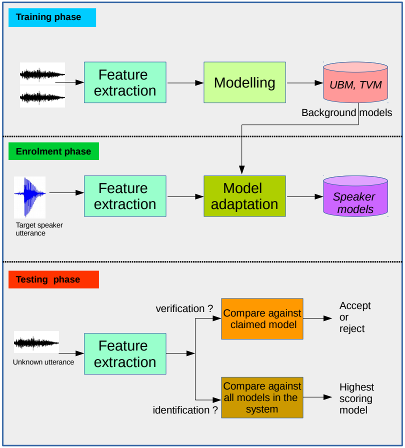


Figure 2 – Phases in Automatic Speaker Recognition Systems

Speaker modeling approaches

- Gaussian mixture models (GMM) [1]
- GMM-Universal background models (GMM-UBM) [1]
- GMM-supervector+SVM [1]
- Joint factor analysis [2]
- i-vectors (state of the art)[2]
- Deep neural networks [3]

1. Tomi Kinnunen and Haizhou Li, "An overview of text-independent speaker recognition: from features to supervectors", Speech communication, 2010.
2. N. Dehak, P.J. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet. "Front-End Factor Analysis for Speaker Verification", IEEE TASLP, 2011.
3. F. Richardson, D. Reynolds, and N. Dehak. "Deep neural network approaches to speaker and language recognition", IEEE Signal Processing Letters, October 2015.

Spoofing voice biometric (ASV) system

- Spoofing vs Anti-spoofing ?
- Spoofing attacks:
  - Impersonation
  - Replay
  - Text-to-Speech
  - Voice conversion

Difficulty level: Spoofer perspective

1. Replay
2. Text-to-Speech
3. Voice conversion
4. Impersonation

Difficulty level: research perspective

1. Text-to-Speech
2. Voice conversion
3. Replay
4. Impersonation

Where do we stand ?

Our main focus

1. Replay
2. Impersonation

Need for ASV anti-spoofing ??

- ASV systems vulnerable to spoofing attacks [1]
- Commercial applications : Adobe TTS, Lyrebird [2,3]

1. Z. Wu, N. Evans, T. Kinnunen, J. Yamagishi, F. Alegre, and H. Li. “Spoofing and countermeasures for speaker verification: a survey”, Speech Communications, 2015.
2. https://lyrebird.ai/
3. https://helpx.adobe.com/audition/using/text-to-speeech.html

ASV Spoofing challenge Overview

ASVSpoof 2015 challenge: 1st edition [1,3]

- Special session at Interspeech 2015.
- Focus on TTS and VC spoofing .
- 16 research teams.
- Text-independent.
- Released ASVspoof 2015 corpus.

BTAS 2016 challenge [2]

- TTS, VC and Replay spoofing.
- 5 research teams.
- Text-independent.
- Released avspoof corpus.

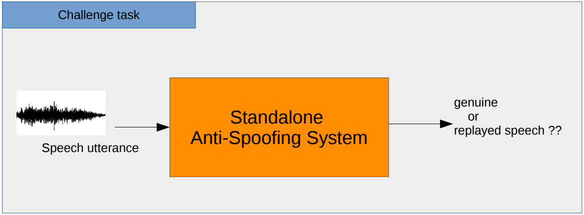
ASVSpoof 2017 challenge: 2nd edition [1,4]

- Special session at Interspeech 2017.
- Focus on Replay spoofing .
- 48 research teams.
- Text-dependent.
- Released ASVspoof 2017 corpus.

Growing interest in the community

1. http://www.asvspoof.org/
2. https://ieee-biometrics.org/btas2016/
3. Zhizheng Wu et. al, "ASVspoof 2015: the First Automatic Speaker Verification Spoofing and Countermeasures Challenge", Interspeech 2015.
4. Tomi Kinnunen et. al, "The ASVspoof 2017 Challenge: Assessing the Limits of Audio Replay Attack Detection in the Wild", Interspeech 2017 (to appear).

ASVSpoof 2017 spoofing challenge



ASVSpoof2017 Dataset

subset	# speakers	# genuine	# spoofed
training	10	1508	1508
development	08	760	950
evaluation	24	1298	12922

Our anti-spoofing system

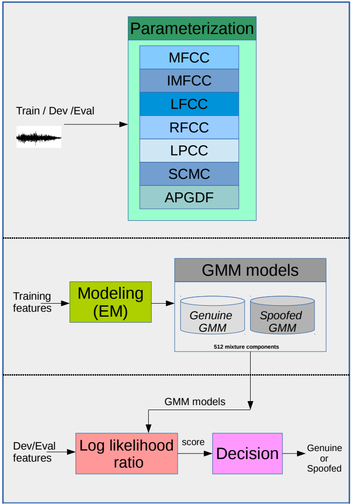


Fig3: Single feature-based anti-spoofing system

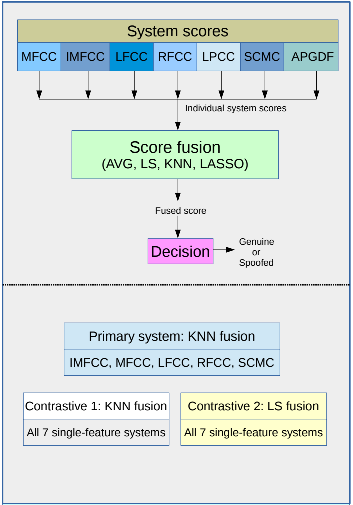


Fig4: Score fusion based anti-spoofing system

Performance

Table 1: Performance, based on equal error rate (EER %), on ASVspoof 2017 development and evaluation data.

System	Development set	Evaluation set
baseline	11.4	30.6
Primary	1.9 ± 0.73	34.78
Contrastive1	2.12 ± 0.76	37.65
Contrastive2	3.25 ± 0.84	36.33

ASVSpoof 2017 Challenge results

Table 2: Top 5 systems of ASVSpoof 2017 replay spoofing challenge [1]

System Name	EER	Description
Baseline	30.6	Based on CQCC 90d
S01	**6.73**	CNN+GMM, iVector+SVM,CNN-RNN; score fusion.
S02	**12.39**	PLP, MFCC and CQCC system fusion.
S03	**14.31**	8 features; GMM and FFNN; fusion.
S04	14.93	6 features; GMM; fusion.
S05	16.35	FBank features; GMM and CTDNN; fusion.

1. Tomi Kinnunen et. al, "The ASVspoof 2017 Challenge: Assessing the Limits of Audio Replay Attack Detection in the Wild", Interspeech 2017 (to appear).

Post-evalution experiments

Table 3: Fused systems obtained after post evaluation. F1-F4 are static+delta+acceleration (SDA) 60d-based score fusion systems. S1-S7 corresponds to MFCC, IMFCC, LFCC, RFCC,LPCC, SCMC and APGDF based systems.

System	Fusion	Dev set	Eval set
F1	S1-S7+B (KNN)	2.76 ± 1.02	33.64
F2	S1-S7+B (AVG)	7.56	31.39
F3	S1-S6+B (AVG)	7.74	30.4
F4	S1-S5+B (AVG)	8.03	**29.17**
F5	S1 (**S**)	4.33	34.3
F6	S1 (**SDA**)	5.44	**30.8**

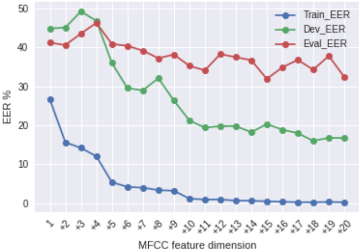
MFCC Vs IMFCC performance

Table 4: Comparing performance of 20 dimensional static MFCC and IMFCC GMM systems trained using 10EM iterations.

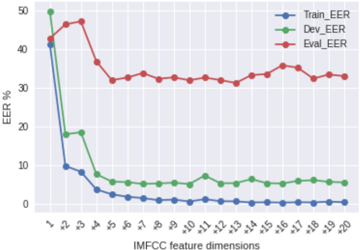
Model order	**Train**		**Dev**		**Eval**	
	MFCC	IMFCC	MFCC	IMFCC	MFCC	IMFCC
512	0.06	0.04	15.6	4.5	35.3	35.2
64	0.19	0.19	14.8	5.03	33.7	34.2
32	0.24	0.51	17.1	5.4	40.4	31.5

Performance on feature dimension

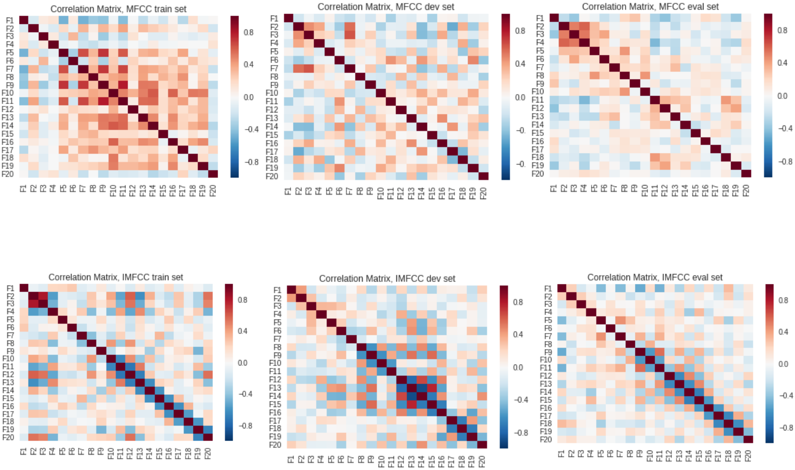
(a) MFCC-based GMM model order = 64.



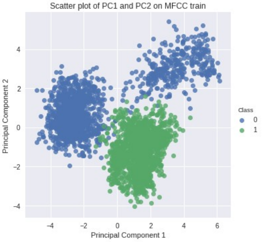
(b) IMFCC-based GMM model order = 32.

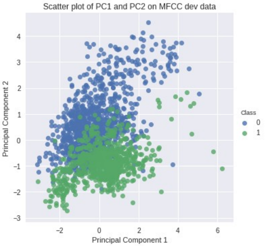


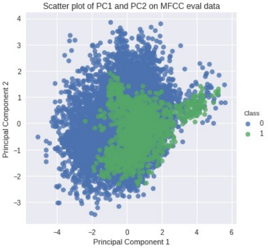
Multivariate analysis: Correlation

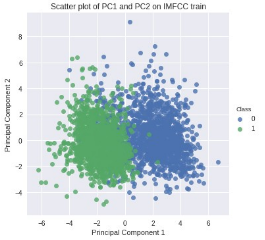


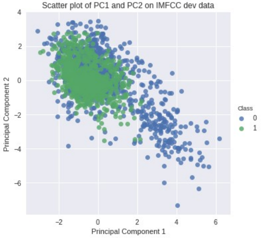
Multivariate analysis: PCA

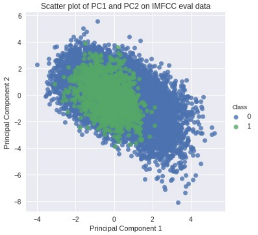












Main progress

1. Database for ASV and spoofing research.
2. Research collaboration: Sheffield University & University of Eastern Finland.
3. Literature review: ASV spoofing.
4. Actively been supervised: 18 supervision logs.
5. Submitted paper in Interspeech-2017.
6. Multi-variate analysis work (going on).

End goals

1. Build speaker models to combat mimicry and replay spoofing attacks.
2. **Alternative applications** of speaker models: spoken language learning, entertainment.
3. Investigating **neural network** approaches to anti-spoofing.