This presentation is the third part of a tutorial presented at INTERSPEECH 2018:

Spoofing attacks in Automatic Speaker Verification: Analysis and Countermeasures

Haizhou Li (National University of Singapore, Singapore), Hemant A. Patil (Dhirubhai Ambani Institute of Information and Communication Technology, India), Nicholas Evans (EURECOM, France)

Abstract: Speech is the most natural means of communication between humans. Speech signals carry various levels of information, such as linguistic content, emotion, the acoustic environment, language, the speaker's identity and their health condition, etc. Automatic speaker recognition technologies aim to verify or identify a speaker using recordings of his/her voice. In practice, automatic speaker verification (ASV) systems should be robust to nuisance variation such as differences in the microphone and transmission channel, intersession variability, acoustic noise, speaker ageing, etc. Significant effort invested over the last three decades has been tremendously successful in developing technologies to compensate for such nuisance variation, thereby improving the reliability of ASV systems in a multitude of diverse application scenarios. In a number of these, specifically those relating to authentication applications, reliability can still be compromised as a result of spoofing attacks whereby fraudsters can gain illegitimate access to protected resources or facilities through the presentation of specially crafted speech signals that reflect the characteristics of another, enrolled person’s voice. ASV systems should be resilient to such malicious spoofing attacks. This tutorial presents a treatment of the issues concerning the robustness and security of an ASV system in the face of spoofing attacks. We also discuss current research trends and progress in developing anti-spoofing countermeasures to protect against attacks derived from voice conversion, speech synthesis, replay, twins (which has more malicious nature in attacking ASV systems and also called as twin’s fraud in biometrics literature) and professional mimics. The tutorial will give an overview of the risk and technological challenges associated with each form of attack in addition to an overview of the two internationally competitive ASVspoof challenges held as special sessions at INTERSPEECH 2015 and INTERSPEECH 2017. The tutorial will conclude with a summary of the current state-of-the-art in the field and a discussion of future research directions.

This PDF contains the full slide set that contains additional slides that were not presented during the tutorial. The reduced set of slides that were presented in Hyderabad are available at http://www.eurecom.fr/~evans/interspeech2018/.
Acknowledgements

Massmiliano Todisco
EURECOM

Tomi H. Kinnunen
UEF, Finland

Junichi Yamagishi
Univ. of Edinburgh, UK
NII, Japan

Hector Delgado
EURECOM

Kong Aik Lee
NEC, Japan

Md Sahidullah
UEF, Finland

Douglas Reynolds
MIT Lincoln Laboratory
USA

Zhizheng Wu
JD.COM, USA

Spoofing attacks in automatic speaker verification: analysis and countermeasures

Overview

1999 2006 2014 2016 2017

- small, purpose collected datasets
- adapted, standard datasets
- EU FP7 TABULA RASA
- TABULA RASA
- 2013 Interspeech special session
- common datasets, metrics, protocols
- ASVspoof initiative
- ASVspoof 2015
- ASVspoof 2017
- RedDots Replayed
- common datasets, physical & local access scenarios, generalisation
Impersonation

- human-altered speech
  - skilled attack dependent on voice similarity
- generally very few speakers
- inconsistent findings
  - human listeners v’s ASV
  - prosody v’s timbre

Replay

- representation of previously recorded, bona fide speech
- small number of speaker, but consistent findings
- countermeasures:
  - audio forensic approaches, i.e. channel effects
  - passive, challenge-response, e.g. prompted-text
### Speech synthesis

- artificial, speaker indicative speech
- large, standard datasets, e.g. WSJ
- significant, universal susceptibility
- countermeasures: phase spectra and prosody
  - encouraging potential

<table>
<thead>
<tr>
<th>Study</th>
<th># target speakers</th>
<th>ASV system</th>
<th>Before spoofing</th>
<th>After spoofing</th>
<th>With CMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lindberg 1999</td>
<td>2</td>
<td>HMM</td>
<td>6%</td>
<td>39%</td>
<td>n/a</td>
</tr>
<tr>
<td>Masuko 1999</td>
<td>20</td>
<td>HMM</td>
<td>0%</td>
<td>70%</td>
<td>n/a</td>
</tr>
<tr>
<td>De Leon 2012</td>
<td>283</td>
<td>GMM-UBM</td>
<td>0%</td>
<td>86%</td>
<td>2.5%</td>
</tr>
<tr>
<td>De Leon 2012</td>
<td>283</td>
<td>SVM</td>
<td>0%</td>
<td>81%</td>
<td>2.5%</td>
</tr>
</tbody>
</table>

### Voice conversion

- large, standard datasets, e.g. NIST SRE
- universal susceptibility
- countermeasures: phase, prosody and dynamics
  - encouraging potential

<table>
<thead>
<tr>
<th>Study</th>
<th># target speakers</th>
<th>ASV system</th>
<th>Before spoofing</th>
<th>After spoofing</th>
<th>With CMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perrot 2005</td>
<td>n/a</td>
<td>GMM-UBM</td>
<td>~16 %</td>
<td>26%</td>
<td>~40 %</td>
</tr>
<tr>
<td>Matrouf 2006</td>
<td>n/a</td>
<td>GMM-UBM</td>
<td>~8 %</td>
<td>~63 %</td>
<td>~100 %</td>
</tr>
<tr>
<td>Kinnunen 2012</td>
<td>504</td>
<td>JFA</td>
<td>3%</td>
<td>8%</td>
<td>17%</td>
</tr>
<tr>
<td>Wu 2012</td>
<td>504</td>
<td>PLDA</td>
<td>3%</td>
<td>11%</td>
<td>41%</td>
</tr>
<tr>
<td>Alegre 2013</td>
<td>298</td>
<td>PLDA</td>
<td>3%</td>
<td>20%</td>
<td>~55 %</td>
</tr>
<tr>
<td>Kons 2013</td>
<td>750</td>
<td>HMM-NAP</td>
<td>1%</td>
<td>3%</td>
<td>36%</td>
</tr>
</tbody>
</table>
Spoofing attacks in automatic speaker verification: analysis and countermeasures

TABULA RASA - EU FP7

- biometrics
  - ICAO and non-ICAO modalities

- objectives:
  - evaluate spoofing vulnerabilities
  - develop countermeasures
  - exploitation and technology transfer
  - dissemination, standards and ethics
Limitations

- different datasets, protocols and metrics
- inappropriate use of prior knowledge
  - spoofing attacks – system
  - countermeasures – spoofing attacks
- different approaches to integration
- different application scenarios:
  - physical / logical access
  - microphone and channel variations
- lagging behind efforts in other biometrics communities

The making of ASVspoof

- Interspeech 2013 special session
  - spoofing and countermeasures for automatic speaker verification
  - 6 papers
- establish a community-driven initiative
  - address limitations
  - promote consideration of spoofing / vulnerabilities
  - encourage greater participation
  - foster advances in countermeasure design
- standard databases, protocols and metrics
Definitions

- a.k.a. presentation attacks (ISO / IEC)
  ISO/IEC 30107-1

- “persons masquerading as others in order to gain illegitimate access to sensitive or protected resources”

- sensor level: before and after microphone
  - a somewhat contentious issue

Use case scenarios

- physical access
  - fixed microphone / consistent channel
  - spoofing: replay

- logical access
  - microphone and channel unpredictable
  - spoofing: synthetic speech and converted voice
## Priorities

<table>
<thead>
<tr>
<th>Spoofing attack</th>
<th>Accessibility</th>
<th>Effectiveness (risk)</th>
<th>Countermeasure availability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Text-independent</td>
<td>Text-dependent</td>
</tr>
<tr>
<td>Impersonation</td>
<td>Low</td>
<td>Low/unknown</td>
<td>Low/unknown</td>
</tr>
<tr>
<td>ASVspoof 2015</td>
<td><strong>High</strong></td>
<td>Low/unknown</td>
<td><strong>Low</strong></td>
</tr>
<tr>
<td>ASVspoof 2017</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speech synthesis</td>
<td>Medium to high</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>ASVspoof 2015</td>
<td><strong>Medium</strong></td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Voice conversion</td>
<td>Medium to high</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>ASVspoof 2019 ?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ASVspoof: guiding principles

- motivation
  - improve the research methodology and generalisation
  - common databases, protocols and metrics: level playing field
  - advance the state of the art in spoofing countermeasures

- isolated spoofing detection, speaker independent

Spoofing detection → Score: high → bona fide
low → spoof

- requires no expertise in automatic speaker verification

ASVspoof 2015

- logical access
- speech synthesis (TTS) and voice conversion (VC)

Train
- ground-truth
- 25 speakers
- 5 TTS and VC attacks

Development
- ground-truth
- 35 speakers
- 5 TTS and VC attacks

Evaluation
- NO ground-truth
- unknown attacks
- 46 speakers
- 10 TTS and VC attacks
ASVspoof 2015 – spoofing attacks

- **S1 – S5**: in the training, development & evaluation sets
  - **S1**: VC - Frame selection
  - **S2**: VC - Slope shifting
  - **S3**: TTS – HTS with 20 adaptation sentences
  - **S4**: TTS – HTS with 40 adaptation sentences
  - **S5**: VC – Festvox (http://festvox.org/)

- **S6 – S10**: Only appear in the evaluation set
  - **S6**: VC – ML-GMM with GV enhancement
  - **S7**: VC – Similar to S6 but using LSP features
  - **S8**: VC – Tensor (eigenvoice)-based approach
  - **S9**: VC – Nonlinear regression (KPLS)
  - **S10**: TTS – MARY TTS unit selection (http://mary.dfki.de/)

ASVspoof 2015 – dimensions

<table>
<thead>
<tr>
<th></th>
<th># utterances</th>
<th>Algorithm</th>
<th>Vocoder</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Train</td>
<td>Dev.</td>
<td>Eval.</td>
</tr>
<tr>
<td>Genuine</td>
<td>3750</td>
<td>3497</td>
<td>9404</td>
</tr>
<tr>
<td>S1</td>
<td>2525</td>
<td>9975</td>
<td>18400</td>
</tr>
<tr>
<td>S2</td>
<td>2525</td>
<td>9975</td>
<td>18400</td>
</tr>
<tr>
<td>S3</td>
<td>2525</td>
<td>9975</td>
<td>18400</td>
</tr>
<tr>
<td>S4</td>
<td>2525</td>
<td>9975</td>
<td>18400</td>
</tr>
<tr>
<td>S5</td>
<td>2525</td>
<td>9975</td>
<td>18400</td>
</tr>
<tr>
<td>S6</td>
<td>0</td>
<td>0</td>
<td>18400</td>
</tr>
<tr>
<td>S7</td>
<td>0</td>
<td>0</td>
<td>18400</td>
</tr>
<tr>
<td>S8</td>
<td>0</td>
<td>0</td>
<td>18400</td>
</tr>
<tr>
<td>S9</td>
<td>0</td>
<td>0</td>
<td>18400</td>
</tr>
<tr>
<td>S10</td>
<td>0</td>
<td>0</td>
<td>18400</td>
</tr>
</tbody>
</table>
ASVspoof 2015 – vulnerabilities

*Baseline*, EER = 2.30
[S1], EER = 40.42
[S2], EER = 2.66
[S3], EER = 40.29
[S4], EER = 43.35
[S5], EER = 46.24
[S6], EER = 44.71
[S7], EER = 29.29
[S8], EER = 36.17
[S9], EER = 33.52
[S10], EER = 51.17

Spoofing attacks in automatic speaker verification: analysis and countermeasures

ASVspoof 2015 – bona fide or spoof?

This slide originally contained audio samples and animations that were played during the tutorial and that were lost during conversion to PDF format.
ASVspoof 2015 – results

<table>
<thead>
<tr>
<th>Team</th>
<th>Known attacks (S1 - S5)</th>
<th>Unknown attacks (S6 - S10)</th>
<th>Average (all)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DA-IICT</td>
<td>0.408</td>
<td>2.013</td>
<td>1.211</td>
</tr>
<tr>
<td>STC</td>
<td>0.008</td>
<td>3.922</td>
<td>1.965</td>
</tr>
<tr>
<td>SJTU</td>
<td>0.058</td>
<td>4.998</td>
<td>2.528</td>
</tr>
<tr>
<td>NTU</td>
<td>0.003</td>
<td>5.231</td>
<td>2.617</td>
</tr>
<tr>
<td>CRIM</td>
<td>0.041</td>
<td>5.347</td>
<td>2.694</td>
</tr>
<tr>
<td>F</td>
<td>0.358</td>
<td>6.078</td>
<td>3.218</td>
</tr>
<tr>
<td>G</td>
<td>0.405</td>
<td>6.247</td>
<td>3.326</td>
</tr>
<tr>
<td>H</td>
<td>0.67</td>
<td>6.041</td>
<td>3.355</td>
</tr>
<tr>
<td>I</td>
<td>0.005</td>
<td>7.447</td>
<td>3.726</td>
</tr>
<tr>
<td>J</td>
<td>0.025</td>
<td>8.168</td>
<td>4.097</td>
</tr>
<tr>
<td>K</td>
<td>0.21</td>
<td>8.883</td>
<td>4.547</td>
</tr>
<tr>
<td>L</td>
<td>0.412</td>
<td>13.026</td>
<td>6.719</td>
</tr>
<tr>
<td>M</td>
<td>8.528</td>
<td>20.253</td>
<td>14.391</td>
</tr>
<tr>
<td>N</td>
<td>7.874</td>
<td>21.262</td>
<td>14.568</td>
</tr>
<tr>
<td>O</td>
<td>17.723</td>
<td>19.929</td>
<td>18.826</td>
</tr>
</tbody>
</table>

28 teams requested data
16 teams submitted results

Best performance overall & for S10
Best performance for S1 – S9

Top-performing systems

- **DA-IICT**
  - cochlear filter cepstral coefficients plus instantaneous frequency (CFCCIF) with MFCC, GMM
  - T. B. Patel and H. A. Patil, Combining evidences from Mel cepstral, cochlear filter cepstral and instantaneous frequency features for detection of natural vs. spoofed speech, Interspeech 2015

- **STC**
  - Mel-frequency principle coefficients, CosPhase principle coefficients, Mel wavelet packet coefficients, MFCC, i-vector

- **SJTU**
  - deep learning, i-vector / PLDA
  - N. Chen, Y. Qian, H. Dinkel, B. Chen and K. Yu, Robust deep features for spoofing detection – the SJTU systems for ASVspoof 2015 challenge, Interspeech 2015

- **NTU**
  - high-resolution phase and magnitude features, MLP


N. Chen, Y. Qian, H. Dinkel, B. Chen and K. Yu, Robust Deep Feature for Spoofing Detection - The SJTU System for ASVspoof 2015 Challenge

A. Janicki, Spoofing Countermeasures Based on Analysis of Linear Prediction Error

Y. Liu, Y. Tian, L. He, J. Liu and M. T. Johnson, Simultaneous Utilization of Spectral Magnitude and Phase Information to Extract Supervectors for Speaker Verification Anti-spoofing

T. B. Patel and H. A. Patil, Combining Evidences from Mel Cepstral, Cochlear Filter Cepstral and Instantaneous Frequency Features for Detection of Natural vs. Spoofed Speech


J. Villalba, A. Miguel, A. Ortega and E. Lleida, Spoofing Detection with DNN and One-class SVM for the ASVspoof 2015 Challenge

L. Wang, Y. Yoshida, Y. Kawakami and S. Nakagawa, Relative phase information for detecting human speech and spoofed speech


 Spoofing attacks in automatic speaker verification: analysis and countermeasures
CQCC results for ASVspoof

<table>
<thead>
<tr>
<th>System</th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>S4</th>
<th>S5</th>
<th>Avg.</th>
<th>S6</th>
<th>S7</th>
<th>S8</th>
<th>S9</th>
<th>S10</th>
<th>Avg.</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFCC-IF</td>
<td>0.101</td>
<td>0.863</td>
<td>0.000</td>
<td>0.000</td>
<td>1.075</td>
<td>0.408</td>
<td>0.846</td>
<td>0.212</td>
<td>0.142</td>
<td>0.346</td>
<td>8.490</td>
<td>2.013</td>
<td>1.211</td>
</tr>
<tr>
<td>i-vector</td>
<td>0.004</td>
<td>0.022</td>
<td>0.000</td>
<td>0.000</td>
<td>0.013</td>
<td>0.008</td>
<td>0.019</td>
<td>0.000</td>
<td>0.015</td>
<td>0.004</td>
<td>19.57</td>
<td>3.922</td>
<td>1.965</td>
</tr>
<tr>
<td>DNN feat.</td>
<td>0.032</td>
<td>0.109</td>
<td>0.032</td>
<td>0.032</td>
<td>0.086</td>
<td>0.058</td>
<td>0.173</td>
<td>0.049</td>
<td>0.121</td>
<td>0.049</td>
<td>24.601</td>
<td>4.998</td>
<td>2.528</td>
</tr>
<tr>
<td>LFCC-DA</td>
<td>0.027</td>
<td>0.408</td>
<td>0.000</td>
<td>0.000</td>
<td>0.114</td>
<td>0.110</td>
<td>0.149</td>
<td>0.011</td>
<td>0.074</td>
<td>0.027</td>
<td>8.185</td>
<td>1.670</td>
<td>0.800</td>
</tr>
<tr>
<td>CQCC-A</td>
<td>0.005</td>
<td>0.106</td>
<td>0.000</td>
<td>0.000</td>
<td>0.130</td>
<td>0.048</td>
<td>0.008</td>
<td>0.064</td>
<td>1.033</td>
<td>0.053</td>
<td>1.065</td>
<td>0.462</td>
<td>0.255</td>
</tr>
</tbody>
</table>

- competitive results for known attacks
- best results for unknown attacks:
  - attack S10 (unit selection): 87% relative improvement
  - overall: 72% relative improvement

Winner of best paper award at the Speaker and Language Recognition Workshop (ODYSSEY) 2016

M. Todisco, H. Deglado and N. Evans, A new feature for automatic speaker verification anti-spoofing: constant Q cepstral coefficients, Speaker Odyssey 2016

M. Todisco, H. Delgado and N. Evans, Constant Q cepstral coefficients: A spoofing countermeasure for automatic speaker verification, Computer Speech & Language, 2017

ASVspoof 2015 – summary

- high-tech attacks, no replay
- isolated spoofing detection – no ASV
- text-independent
- greatest effort: features
- mostly simple classifiers
- lack of generalisation
- post-evaluation improvements

RedDots Replayed – 2016

RedDots Replayed

- most damaging form of spoofing attack
  - speech synthesis and voice conversion?

- most prolific form of spoofing attack
  - replay?

- ASVspoof 2015
  - logical access
  - text-independent speaker recognition

- RedDots Replayed
  - physical access
  - text-dependent speaker recognition
speaker recognition

objectives:
- spoofing countermeasures
- environmental robustness
- commercial-grade and hybrid ASV
- scalable, trusted biometric authentication service

Replay scenarios and countermeasures

- phrase prompting with utterance verification
  - did the user speak the prompted text?
  
  can be circumvented using voice conversion

- audio fingerprinting
  - do I know this recording
  
  dynamically increasing database size

- speaker-independent replay detection
  - is this recording authentic or replayed one?
  
  most general - but can it be done?
A realistic scenario

Worst case / simulated scenario

Data collection

Allows the use of existing speech corpora to create replay attacks


Crowd-sourced replay attacks

Re-recording of existing RedDots corpora

- text-dependent automatic speaker verification
- collected by volunteers (ASV researchers)
- various Android devices, speakers, accents

https://sites.google.com/site/thereddotsproject/

Example replay configurations

- Smartphone $\rightarrow$ Smartphone
- High-quality loudspeaker $\rightarrow$ smartphone, anechoic room
- Headphones $\rightarrow$ PC mic
- High-quality loudspeaker $\rightarrow$ high-quality mic
- Laptop line-out $\rightarrow$ PC line-in using a cable

Results

<table>
<thead>
<tr>
<th>Type of impostor</th>
<th>GMM-UBM</th>
<th>i-vector (cosine)</th>
<th>i-vector (PLDA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zero effort</td>
<td>2.50</td>
<td>6.64</td>
<td>5.23</td>
</tr>
<tr>
<td>Replay</td>
<td>23.18</td>
<td>26.63</td>
<td>24.85</td>
</tr>
</tbody>
</table>

Replay attack detection results (EER %), Gaussian mixture model classifier

<table>
<thead>
<tr>
<th>Front-end feature</th>
<th>Controlled</th>
<th>Variable</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>LFCC 20-da</td>
<td>5.88</td>
<td>4.43</td>
<td>5.11</td>
</tr>
<tr>
<td>CQCC 20-a</td>
<td>2.77</td>
<td>3.50</td>
<td>3.27</td>
</tr>
</tbody>
</table>

Speakers: 62
Data: 3854 genuine, 52944 spoofed

ASVspoof 2017

- physical access and replay spoofing attacks

**Train**
- ground-truth
- 10 speakers
- 3 replay configurations

**Development**
- ground-truth
- 8 speakers
- 10 replay configurations

**Evaluation**
- NO ground-truth
- unknown attacks
- 24 speakers
- 110 replay configurations
### ASVspoof 2017 – dimensions

<table>
<thead>
<tr>
<th>Subset</th>
<th># Speakers</th>
<th># Replay sessions</th>
<th># Replay configs</th>
<th># utterances</th>
<th>Bona fide</th>
<th>Replay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>10</td>
<td>6</td>
<td>3</td>
<td>1508</td>
<td>1508</td>
<td></td>
</tr>
<tr>
<td>Development</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>760</td>
<td>950</td>
<td></td>
</tr>
<tr>
<td>Evaluation</td>
<td>24</td>
<td>161</td>
<td>110</td>
<td>1298</td>
<td>12008</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>42</td>
<td>177</td>
<td>123</td>
<td>3566</td>
<td>14466</td>
<td></td>
</tr>
</tbody>
</table>

Largest, most diverse replay dataset

### ASVspoof 2017 – vulnerabilities

Bona fide vs. zero-effort impostors
EER = 1.8%

Bona fide vs. replay spoofs
EER = 31.5%

GMM-UBM ASV system

ASVspoof 2017 evaluation set
ASVspoof 2017 – bona fide or replay?

This slide originally contained audio samples and animations that were played during the tutorial and that were lost during conversion to PDF format.

ASVspoof 2017 – results

- 113 database download requests
- 21 of 49 submissions outperformed the baseline (train + dev)
- S01: > 70% relative improvement w.r.t baseline B01
## ASVspoof – top 10 submissions

<table>
<thead>
<tr>
<th>ID</th>
<th>EER</th>
<th>Features</th>
<th>Post-proc.</th>
<th>Classifiers</th>
<th>Fusion</th>
<th>#Subs.</th>
<th>Training</th>
</tr>
</thead>
<tbody>
<tr>
<td>S01</td>
<td>6.73</td>
<td>Log-power Spectrum, LPCC</td>
<td>MVN</td>
<td>CNN, GMM, TV, RNN</td>
<td>Score</td>
<td>3</td>
<td>T</td>
</tr>
<tr>
<td>S02</td>
<td>12.34</td>
<td>CQCC, MFCC, PLP</td>
<td>WMVN</td>
<td>GMM-UBM, TV-PLDA, GSV-SVM, GSV-GBDT, GSV-RF</td>
<td>Score -</td>
<td>T</td>
<td></td>
</tr>
<tr>
<td>S03</td>
<td>14.03</td>
<td>MFCC, IMFCC, RFCC, LFCC, PLP, CQCC, SCMC, SSFC</td>
<td>-</td>
<td>GMM, FF-ANN</td>
<td>Score</td>
<td>18</td>
<td>T+D</td>
</tr>
<tr>
<td>S04</td>
<td>14.66</td>
<td>RFCC, MFCC, IMFCC, LFCC, SSFC, SCMC</td>
<td>-</td>
<td>GMM</td>
<td>Score</td>
<td>12</td>
<td>T+D</td>
</tr>
<tr>
<td>S05</td>
<td>15.97</td>
<td>Linear filterbank feature</td>
<td>MN</td>
<td>GMM, CT-DNN</td>
<td>Score</td>
<td>2</td>
<td>T</td>
</tr>
<tr>
<td>S06</td>
<td>17.62</td>
<td>CQCC, IMFCC, SCMC, Phrase one-hot encoding</td>
<td>MN</td>
<td>GMM</td>
<td>Score</td>
<td>4</td>
<td>T+D</td>
</tr>
<tr>
<td>S07</td>
<td>18.14</td>
<td>HPCC, CQCC</td>
<td>MVN</td>
<td>GMM, CNN, SVM</td>
<td>Score</td>
<td>2</td>
<td>T+D</td>
</tr>
<tr>
<td>S08</td>
<td>18.32</td>
<td>IFCC, CEFCIF, Prosody</td>
<td>-</td>
<td>GMM</td>
<td>Score</td>
<td>3</td>
<td>T</td>
</tr>
<tr>
<td>S10</td>
<td>20.32</td>
<td>CQCC</td>
<td>-</td>
<td>ResNet</td>
<td>None</td>
<td>1</td>
<td>T</td>
</tr>
<tr>
<td>S09</td>
<td>20.57</td>
<td>SFFCC</td>
<td>-</td>
<td>GMM</td>
<td>None</td>
<td>1</td>
<td>T</td>
</tr>
<tr>
<td>D01</td>
<td>7.00</td>
<td>MFCC, CQCC, WT</td>
<td>MVN</td>
<td>GMM, TV-SVM</td>
<td>Score</td>
<td>26</td>
<td>T+D</td>
</tr>
</tbody>
</table>

### Spoofing attacks in automatic speaker verification: analysis and countermeasures

- M. Wilkowski, S. Kacprzak, P. Żelasko, K. Kowalczyk and J. Galka, *Audio Replay Attack Detection Using High-Frequency Features*
- X. Wang, Yanhong Xiao and Xuan Zhu, *Feature Selection Based on CQCCs for Automatic Speaker Verification Spoofing*
- Z. Ji, Z.-Y. Li, P. Li, M. An, S. Gao, D. Wu and F. Zhao, *Ensemble Learning for Countermeasure of Audio Replay Spoofing Attack in ASVspoof2017*
- Z. Chen, Z. Xie, W. Zhang and X. Xu, *ResNet and Model Fusion for Automatic Spoofing Detection*
ASVspoof 2017 – summary

- alignment to text-dependent ASV community
- successful crowdsourcing approach to replay data collection
  - heterogeneous / in-the-wild attacks
- difficulty in analysing results
  - more difficult c.f. ASVspoof 2015
  - generalisation still lacking
- top-ranked system
  - ~70% relative improvement w.r.t. the baseline system
  - fusion of only 3 subsystems
- encouraging performance
  - high detection performance for high quality attacks

ASVspoof 2017 v2.0
ASVspoof 2017 v2.0

- data patching
  - RedDots characteristics
  - results / attention mechanisms

- meta data
  - more meaningful replay characterisations
  - improved analysis

- baseline enhancements
  - log energy
  - CMV normalisation

B. Chettri and B. L. Sturm, A deeper look at Gaussian mixture model based anti-spoofing systems, ICASSP 2018

ASVspoof 2017 – acoustic environments

<table>
<thead>
<tr>
<th>ID</th>
<th>Environment</th>
<th>ID</th>
<th>Environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>E01</td>
<td>Anechoic room</td>
<td>E14</td>
<td>Office 02</td>
</tr>
<tr>
<td>E02</td>
<td>Balcony 01</td>
<td>E15</td>
<td>Office 03</td>
</tr>
<tr>
<td>E03</td>
<td>Balcony 02</td>
<td>E16</td>
<td>Office 04</td>
</tr>
<tr>
<td>E04</td>
<td>Home 07</td>
<td>E17</td>
<td>Office 05</td>
</tr>
<tr>
<td>E05</td>
<td>Home 08</td>
<td>E18</td>
<td>Office 06</td>
</tr>
<tr>
<td>E06</td>
<td>Canine</td>
<td>E19</td>
<td>Office 07</td>
</tr>
<tr>
<td>E07</td>
<td>Home 01</td>
<td>E20</td>
<td>Office 08</td>
</tr>
<tr>
<td>E08</td>
<td>Home 02</td>
<td>E21</td>
<td>Office 09</td>
</tr>
<tr>
<td>E09</td>
<td>Home 03</td>
<td>E22</td>
<td>Office 10</td>
</tr>
<tr>
<td>E10</td>
<td>Home 04</td>
<td>E23</td>
<td>Studio</td>
</tr>
<tr>
<td>E11</td>
<td>Home 05</td>
<td>E24</td>
<td>Analog wire 01</td>
</tr>
<tr>
<td>E12</td>
<td>Home 06</td>
<td>E25</td>
<td>Analog wire 02</td>
</tr>
<tr>
<td>E13</td>
<td>Office 01</td>
<td>E26</td>
<td>Analog wire 03</td>
</tr>
</tbody>
</table>

Noisy environments: canteen (bubble noise), balcony (street noise) ...

No-treated environments: office, home, living room (silent, TV on ...)

Acoustically treated environments (studio, anechoic room) and analog wires

Generic IDs split into specific IDS

Example: Home (generic) → Home 01, Home 02, ...
### ASVspoof 2017 – playback devices

<table>
<thead>
<tr>
<th>ID</th>
<th>Playback device</th>
</tr>
</thead>
<tbody>
<tr>
<td>P01</td>
<td>All-in-one PC speakers</td>
</tr>
<tr>
<td>P02</td>
<td>Creative A90 speakers</td>
</tr>
<tr>
<td>P03</td>
<td>Genelec 8030C studio monitor</td>
</tr>
<tr>
<td>P04</td>
<td>Genelec 8030C studio monitor (2 speakers)</td>
</tr>
<tr>
<td>P05</td>
<td>Beyerdynamic DT 770 PRO headphones</td>
</tr>
<tr>
<td>P06</td>
<td>Dell laptop internal speakers</td>
</tr>
<tr>
<td>P07</td>
<td>Dynaudio RN5A speaker</td>
</tr>
<tr>
<td>P08</td>
<td>HP Laptop internal speakers</td>
</tr>
<tr>
<td>P09</td>
<td>VIFA M10HD-39-18 speaker</td>
</tr>
<tr>
<td>P10</td>
<td>ACER netbook internal speakers</td>
</tr>
<tr>
<td>P11</td>
<td>B&amp;O Aquaris M5 smartphone</td>
</tr>
<tr>
<td>P12</td>
<td>Logitech low-quality speakers</td>
</tr>
<tr>
<td>P13</td>
<td>Desktop PC line output</td>
</tr>
<tr>
<td>P14</td>
<td>Lørger LCS-1050 speakers</td>
</tr>
<tr>
<td>P15</td>
<td>Edirol MA-15D studio monitor</td>
</tr>
<tr>
<td>P16</td>
<td>Lenovo K100 3600G4i tablet</td>
</tr>
<tr>
<td>P17</td>
<td>Logitech S120 multimedia speakers</td>
</tr>
<tr>
<td>P18</td>
<td>MacBook pro internal speakers</td>
</tr>
<tr>
<td>P19</td>
<td>Alesis Lancing USB MIDI 227 portable speaker</td>
</tr>
<tr>
<td>P20</td>
<td>Samsung GT-I9100 smartphone</td>
</tr>
<tr>
<td>P21</td>
<td>Samsung GT-P6800 tablet</td>
</tr>
<tr>
<td>P22</td>
<td>Behringer Truth B2000A studio monitor</td>
</tr>
<tr>
<td>P23</td>
<td>Focusrite Scarlet 2i2 audio interface line output</td>
</tr>
<tr>
<td>P24</td>
<td>Focusrite Scarlet 2i4 audio interface line output</td>
</tr>
<tr>
<td>P25</td>
<td>Genelec 6330A studio monitor</td>
</tr>
<tr>
<td>P26</td>
<td>AKG K242HD Headset</td>
</tr>
</tbody>
</table>

**Devices with small loudspeakers: laptops / smartphones / tablets**

**Larger-size, consumer loudspeakers (desktop speakers)**

**Studio-quality loudspeakers / headphones / analog outputs**

---

### ASVspoof 2017 – recording devices

<table>
<thead>
<tr>
<th>ID</th>
<th>Recording device</th>
</tr>
</thead>
<tbody>
<tr>
<td>R01</td>
<td>Zoom H6 handy recorder</td>
</tr>
<tr>
<td>R02</td>
<td>B&amp;O Aquaris M5 smartphone</td>
</tr>
<tr>
<td>R03</td>
<td>Low-quality headset</td>
</tr>
<tr>
<td>R04</td>
<td>Nokia Lumia 635 smartphone</td>
</tr>
<tr>
<td>R05</td>
<td>Rode NT2 microphone</td>
</tr>
<tr>
<td>R06</td>
<td>Rode smartlav+ microphone</td>
</tr>
<tr>
<td>R07</td>
<td>Samsung Galaxy S7 smartphone</td>
</tr>
<tr>
<td>R08</td>
<td>Desktop PC microphone input</td>
</tr>
<tr>
<td>R09</td>
<td>Zoom H6 recorder with Behringer ECM8000 mic.</td>
</tr>
<tr>
<td>R10</td>
<td>Zoom H6 recorder with MSI-H microphone</td>
</tr>
<tr>
<td>R11</td>
<td>Zoom H6 recorder with XY microphone</td>
</tr>
<tr>
<td>R12</td>
<td>iPhone 5c smartphone</td>
</tr>
<tr>
<td>R13</td>
<td>iPhone 7 plus smartphone</td>
</tr>
<tr>
<td>R14</td>
<td>iPhone 4 smartphone</td>
</tr>
<tr>
<td>R15</td>
<td>Logitech C920 webcam</td>
</tr>
<tr>
<td>R16</td>
<td>miniDSP UMIX-1 microphone</td>
</tr>
<tr>
<td>R17</td>
<td>Samsung Galaxy Trend 2 smartphone</td>
</tr>
<tr>
<td>R18</td>
<td>Samsung GT-I9100 smartphone</td>
</tr>
<tr>
<td>R19</td>
<td>Samsung GT-P6800 tablet</td>
</tr>
<tr>
<td>R20</td>
<td>Samsung Trend 2 smartphone</td>
</tr>
<tr>
<td>R21</td>
<td>AKG C3000 microphone</td>
</tr>
<tr>
<td>R22</td>
<td>SE electronic 2200a microphone</td>
</tr>
<tr>
<td>R23</td>
<td>Focusrite Scarlet 2i2 interface line output</td>
</tr>
<tr>
<td>R24</td>
<td>Focusrite Scarlet 2i4 interface line output</td>
</tr>
<tr>
<td>R25</td>
<td>Zoom HD1 handy recorder</td>
</tr>
</tbody>
</table>

**Devices with small microphones: smartphones / tablets**

**Headset / webcam microphones**

**Studio-quality microphones / hand-held recorders / analog inputs**

---
ASVspoof 2017 – replay configurations

RedDots
playback device

acoustic
environment

RedDotsReplayed
recording device

Subset | # Speakers | # Replay sessions | # Replay configs | # utterances | Bona fide | Replay |
-------|------------|------------------|------------------|--------------|----------|--------|
Training | 10 | 6 | 3 | 1507 | 1507 |
Development | 8 | 10 | 10 | 760 | 950 |
Evaluation | 24 | 161 | 110 | 1298 | 12008 |
Total | 42 | 177 | 123 | 3566 | 14466 |

ASVspoof 2017 – RC impact

- EER of a GMM-UBM ASV system for each replay configuration (RC)
  - E: acoustic environment
  - P: playback device
  - R: recording device
  - low threat
  - medium threat
  - high threat

Spoofing attacks in automatic speaker verification: analysis and countermeasures
ASVspoof 2017 – frontend enhancements

Log-energy coefficients
- replay introduces non-linearities which may affect within utterance energy dynamics
- log-energy calculated in frequency domain over CQT spectrogram

\[ \log E(n) = \log \left( \sum_{k=1}^{K} |X_{CQ}(k, n)|^2 \right) - \log(K) \]

Cepstral mean and variance normalisation
- typically used to remove unwanted channel effects
- counterintuitive, but beneficial
- channel variation in bona fide data
- channel compensation may help to learn what really is distinctive of genuine / replayed signals


ASVspoof 2017 – performance

Spoofing detection performance (EER, %)

<table>
<thead>
<tr>
<th></th>
<th>T</th>
<th>D</th>
<th>T</th>
<th>D</th>
<th>T+D</th>
<th>T</th>
<th>D</th>
<th>T</th>
<th>D</th>
<th>T+D</th>
</tr>
</thead>
<tbody>
<tr>
<td>training on</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>testing on</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feat. config.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no normalisation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CMVN</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GMM</td>
<td>19-SDA</td>
<td>11.69</td>
<td>1.36</td>
<td>30.79</td>
<td>25.33</td>
<td>23.97</td>
<td>13.31</td>
<td>8.49</td>
<td>19.74</td>
<td>16.89</td>
</tr>
<tr>
<td></td>
<td>19E-SDA</td>
<td>10.37</td>
<td>1.37</td>
<td>34.95</td>
<td>26.3</td>
<td>29.31</td>
<td>9.06</td>
<td>5.64</td>
<td>13.74</td>
<td>14.77</td>
</tr>
<tr>
<td>i-vector</td>
<td>19-SDA</td>
<td>4.43</td>
<td>1.23</td>
<td>17.82</td>
<td>18.81</td>
<td>18.69</td>
<td>11.61</td>
<td>8.74</td>
<td>16.61</td>
<td>15.08</td>
</tr>
</tbody>
</table>

- T Train
- D Development
- E Evaluation
- CMVN

No normalisation:
- energy coefficient decreases performance
- i-vector outperforms GMM by a large margin

CMVN
- energy coefficient increases performance
- GMM slightly outperforms i-vector

Spoofing attacks in automatic speaker verification: analysis and countermeasures
ASVspoof 2017 – RC analysis

- no correlation between ASV degradation and replay detection performance
- inconsistent performance for GMM and i-vector CMs

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>E: acoustic environment</td>
<td>16.68</td>
<td>18.73</td>
<td>21.86</td>
</tr>
<tr>
<td>P: playback device</td>
<td>16.64</td>
<td>16.44</td>
<td>18.37</td>
</tr>
<tr>
<td>R: recording device</td>
<td>10.80</td>
<td>15.69</td>
<td>17.77</td>
</tr>
</tbody>
</table>

Acoustic environment seems to have a greater impact on performance.

ASVspoof 2017 v2.0 – summary

- promising performance even for worst case scenario
- observations
  - cues from higher frequencies
  - voice activity detection detrimental to performance
  - variability between different solutions across different RCs
  - log energy and CMVN helpful
- uncontrolled data collection
  - analysis difficult
  - controlled data collection / simulation needed in future
ASVspoof 2019

- logical access AND physical access scenarios
  - state-of-the-art synthetic speech, converted voice and replay
- controlled setup
- based upon VCTK
  - same as ASVspoof 2015
- protocol enhancements
- impact upon ASV
  - metrics and integration
Metrics and integration

- limitations of the previous work
  - independent spoofing countermeasures
  - does not reflect impact upon ASV

- vision for the future
  - reflect integrated systems
  - backward compatibility with standalone assessment
  - allows the specification of application costs and priors
  - facilitates unified comparison of
    - ASV without countermeasure
    - ASV with perfect countermeasure
    - perfect ASV system with countermeasure
  - metric that is easy to understand and use

Integration

Spoofing detection

Score: high → bona fide
low → spoof

two systems – different objectives – how many user classes?

Speaker verification

Score: high → target
low → impostor

claimed ID
ASV trials

πₜₐʳ

Speaker verification

Score

Accept

Reject

pₕₐₗₜₐₜₐₚₚₚₜₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚportioned

CM trials

πₜₐʳ

Bona fide speech

Speaker verification

Score

Accept

Reject

pₕₐₗₜₐₜₐₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚₚportioned

is the speech sample bona fide?

πₜₐₕₙₚₚₚₚₚportioned

Spoofed speech

Spoofing detection

Score

Reject

Accept

pₕₐₗₜₐₜₐₚₚₚportioned

is the person who they say they are?
**Integrated system trials**

- \( \pi_{\text{tar}} \)
- \( \pi_{\text{non}} \)
- \( \pi_{\text{spoof}} \)

Claimed ID: Haizhou

How to integrate?

Integrated system → Score → Reject/Accept

How to assess?

Three user classes!

<table>
<thead>
<tr>
<th>Trial</th>
<th>tar</th>
<th>non</th>
<th>spoof</th>
</tr>
</thead>
<tbody>
<tr>
<td>CM</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASV</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>ASV/CM</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
</tr>
</tbody>
</table>

**Architectures**

- **cascaded combination**
  - Spoofing detection → Speaker verification
    - Claimed ID
  - Parallel combination
  - Spoofing detection → Speaker verification
    - Claimed ID

Spoofing attacks in automatic speaker verification: analysis and countermeasures
Integrated system errors

Spoofing detection

Spoof accepted

non-target accepted

Speaker verification

spoof accepted

bona fide target rejected

bona fide target rejected

Four possible errors

ASV system rejects target

\[ P_a(s, t) = (1 - P_{\text{miss}}^{\text{cm}}(s)) \times P_{\text{miss}}^{\text{asv}}(t) \]

CM falsely accepts spoof

\[ P_c(s, t) = P_{\text{fa}}^{\text{cm}}(s) \times (1 - P_{\text{miss}}^{\text{asv}}(t)) \]

ASV system accepts non-target

\[ P_b(s, t) = (1 - P_{\text{miss}}^{\text{cm}}(s)) \times P_{\text{fa}}^{\text{asv}}(t) \]

CM rejects target

\[ P_d(s) = P_{\text{miss}}^{\text{cm}}(s) \]

Spoofing attacks in automatic speaker verification: analysis and countermeasures

The traditional NIST-defined detection cost function at ASV threshold \( t \)

\[ DCF(t) = C_{\text{miss}}^{\text{asv}} \pi_{\text{tar}} P_{\text{miss}}^{\text{asv}}(t) + C_{\text{fa}}^{\text{asv}} \pi_{\text{non}} P_{\text{fa}}^{\text{asv}}(t) \]

\[ \pi_{\text{tar}} + \pi_{\text{non}} = 1 \]

The tandem detection cost function at ASV threshold \( t \) and CM threshold \( s \)

\[ t-DCF(s, t) = C_{\text{miss}}^{\text{asv}} \pi_{\text{tar}} P_{a}(s, t) + C_{\text{fa}}^{\text{asv}} \pi_{\text{non}} P_{b}(s, t) \]

\[ + C_{\text{fa}}^{\text{cm}} \pi_{\text{spooof}} P_{c}(s, t) + C_{\text{miss}}^{\text{cm}} \pi_{\text{tar}} P_{d}(s) \]

\[ \pi_{\text{tar}} + \pi_{\text{non}} + \pi_{\text{spooof}} = 1 \]

t-DCF: a detection cost function for the tandem assessment of spoofing countermeasures and automatic speaker verification, 
ODYSSEY 2018
Alt. integration

Spoofing detection

Speaker verification


Spoofing attacks in automatic speaker verification: analysis and countermeasures

05/09/2018
## ASVspoof 2019 – tentative schedule

<table>
<thead>
<tr>
<th>Category</th>
<th>Timeline</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development data</td>
<td>late November to early December</td>
</tr>
<tr>
<td>Evaluation data</td>
<td>early-to-mid February</td>
</tr>
<tr>
<td>Score submission</td>
<td>+1 week</td>
</tr>
<tr>
<td>Results</td>
<td>end of February</td>
</tr>
<tr>
<td>Interspeech deadline</td>
<td>29th March</td>
</tr>
</tbody>
</table>

Stay tuned at:  
[http://www.asvspoof.org](http://www.asvspoof.org)

---

Wrap up
Overview

1999 2006 2014 2016 2017

- small, purpose collected datasets
- adapted, standard datasets
- common datasets, metrics, protocols
- 2013 Interspeech special session
- ASVspoof initiative
- ASVspoof 2015
- RedDots Replayed
- ASVspoof 2017
- H2020 OCTAVE
- common datasets, physical & local access scenarios, generalisation
- ASVspoof 2019

Concluding remarks

- previously behind other biometrics communities
  - ASVspoof now among most successful initiative

- ASV systems are vulnerable to spoofing attacks
  - great potential for spoofing countermeasures

- transition from features to classifiers and end-to-end architectures

- evaluation extremely challenging
  - generalisation
  - additive and convolutional noise

- a continuous arms race – a great research topic
  - ASVspoof 2019 !!
New publications

- **INTERSPEECH 2018 – oral session Monday 3rd 15.00, Hall 2: 15:00**

- **INTERSPEECH 2018 – poster session Monday 3rd 16.30 Hall 4-6: Poster 2**
  - Y. Zhao, R. Togneri and V. Sreeram, Spoofing Detection Using Adaptive Weighting Framework and Clustering Analysis
  - S. Jell, S. Kalta, S. R. M. Prasanna and R. Sinra, Exploration of Compressed EPR Features for Replay Attack Detection
  - M. Kambile, H. Tak and H. Patil, Effectiveness of Speech Demodulation-Based Features for Replay Detecto
  - J. Yang, C. You and O. Hu, Feature with Complementarity of Statistics and Principal Information for Spoofing Detection
  - D. Li, L. Wang, J. Dang, M. Liu, Z. Ou, S. Nakagawa, Huijian Guan and Xiangang Li, Multiple Phase Information Combination for Replay Attacks Detection
  - B. Wickramasinghe, S. Irtza, E. Ambikairajah and J. Epps, Frequency Domain Linear Prediction Features for Replay Spoofing Attack Detection
  - F. Tom, M. Jain and P. Jey, End-To-End Audio Replay Attack Detection using Deep Convolutional Networks with Attention
  - Baranya M. S. and H. Murthy, Decision-level Feature Switching as a Paradigm for Replay Attack Detection
  - G. Suthokumar, V. Sethu, C. Wijesriyak and E. Ambikairajah, Modulation Dynamic Features for the Detection of Replay Attacks

- **INTERSPEECH 2018 – poster session Monday 3rd 16.30 Hall 4-6: Poster 3**
  - H. Tak and H. Patil, Novel Linear Frequency Residual Cepstral Features for Replay Attack Detection
  - M. Singh and D. Pati, Linear Prediction Residual Based Short-term Cepstral Features for Replay Attacks Detection

- **Handbook of Biometric Anti-spoofing 2nd edition, Springer 2018**

Speaker recognition software

- **ALIZE 3.0**
  http://www1.i2r.a-star.edu.sg/~alarcher/Softwares.html

- **SPEAR Toolkit (based on BOB)**

- **MSRidentity Toolbox**

- **Kaldi**
  https://github.com/kaldi-asr/kaldi

- **Sidekit**
  http://www-lium.univ-lemans.fr/sidekit/
Databases

- **ASVspoof**
  http://www.asvspoof.org

- **NIST speaker recognition evaluation corpora available from Linguistic Data Consortium**
  https://www.ldc.upenn.edu/

- **RSR2015** [Larcher et al, Interspeech ‘12]

- **RedDots** [Lee et al, Interspeech ‘15]
  https://sites.google.com/site/thereddotsproject/reddots-challenge

- **AVspoof** [Ergünay, BTAS, ‘15]
  https://www.idiap.ch/dataset/avspoof

Spoofing countermeasures

- **Matlab implementation of LFCC, MFCC, IMFCC features extraction from UEF:**
  http://cs.joensuu.fi/~sahid/codes/AntiSpoofing_Features.zip

- **Matlab implementation of CQCC feature extraction from EURECOM:**
  http://audio.eurecom.fr/content/software
Metrics

- Matlab implementation of t-DCF
  http://www.asvs spoof.org/data2017/tDCF_v0.1.zip

Thank you for listening

Haizhou Li  Hemant A. Patil  Nicholas Evans