INTERSPEECH 2018
SEPTEMBER 2-6 | HYDERABAD, INDIA

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/.

ASVspoof

Nicholas Evans

EURECOM
Sophia Antipolis

Updates available from
http://www.eurecom.fr/~evans/interspeech2018/

Acknowledgements

Massmiliano Todisco

EURECOM

Tomi H. Kinnunen

UEF, Finland

Hector Delgado

EURECOM

Junichi Yamagishi

Univ. of Edinburgh, UK
NII, Japan

Md Sahidullah

UEF, Finland

Kong Aik Lee

NEC, Japan

Zhizheng Wu

JD.COM, USA

Douglas Reynolds

MIT Lincoln Laboratory
USA

Spoofing attacks in automatic speaker verification: analysis and countermeasures 115

Overview

1999 2006 2014 2016 2017

small, purpose collected datasets

EU FP7 TABULA RASA

adapted, standard datasets

2013 Interspeech special session

common datasets, metrics, protocols

ASVspoof intiative

ASVspoof 2015

RedDots Replayed

ASVspoof 2017

H2020 OCTAVE

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

					FAR or IER	
Study	# target speakers	# impersonators	ASV system	Feature	Before spoofing	After spoofing
Lau 2004	6	2	GMM-UBM	MFCCs	~0 %	30 ~ 35 %
Lau 2005	4	6	GMM-UBM	MFCCs	~0 %	10 ~ 60 %
Farrus 2010	5	2	k-NN	Prosodic features	5 % (IER)	22 % (IER)
Hautamäki 2013	5	1	i-vector	MFCCs	9 %	12 %

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

			Before spoofing	After spoofing		With countermeasures	
Study	# target speakers	ASV system	EER/FAR	EER	FAR	EER	FAR
Lindberg 1999	2	Text-Dependent HMM	1 ~ 6 %	27 ~ 70 %	90 ~ 100 %	n/a	n/a
Villalba 2011	5	JFA	1%	~ 20 %	68%	0 ~ 14 %	0 ~ 17 %
Wang 2011	13	GMM-UBM	n/a	40%	n/a	10%	n/a

Speech synthesis

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

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

Voice conversion

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

			Before spoofing	After spoofing		With CMs
Study	# target speakers	ASV system	EER/FAR	EER	FAR	FAR
Perrot 2005	n/a	GMM-UBM	~16 %	26%	~40 %	n/a
Matrouf 2006	n/a	GMM-UBM	~8 %	~63 %	~100 %	n/a
Kinnunen 2012	504	JFA	3%	8%	17%	n/a
Wu 2012	504	PLDA	3%	11%	41%	2%
Alegre 2013	298	PLDA	3%	20%	~55 %	4%
Kons 2013	750	HMM-NAP	1%	3%	36%	n/a

- Y. Lau, D. Tran and M. Wagner, *Testing voice mimicry with the YOHO speaker verification corpus*, Knowledge-Based Intelligent Information and Engineering Systems, 2005
- Y. Lau, M. Wagner and D. Tran, Vulnerability of speaker verification to voice mimicking, Proc. Int. Symposium on Intelligent Multimedia, Video and Speech Processing, 2004
- R. G. Hautamäki, T. Kinnunen, V.Hautamäki and T. Leino, A. M. Laukkanen, *i-vectors meet imitators: on vulnerability of speaker verification systems against voice mimicry*, Interspeech 2013
- J. Lindberg, M. Blomberg, et al., *Vulnerability in speaker verification-a study of technical impostor techniques*, Eurospeech 1999
- J. Villalba and E. Lleida, *Detecting replay attacks from far-field recordings on speaker verification systems*, Biometrics and ID Management. Springer. Lecture Notes in Computer Science, 2011
- Z. F. Wang, G. Wei and Q. H. He, *Channel pattern noise based playback attack detection algorithm for speaker recognition*, ICMLC 2011
- T. Masuko, T. Hitotsumatsu, K. Tokuda and T. Kobayashi, *On the security of HMM-based speaker verification systems against imposture using synthetic speech*, Eurospeech 1999
- P. L. De Leon, M. Pucher, J. Yamagishi, I. Hernaez and I. Saratxaga, *Evaluation of speaker verification security and detection of HMM-based synthetic speech.* IEEE TASLP 2012
- P. L. De Leon, B. Stewart and J. Yamagishi, *Synthetic speech discrimination using pitch pattern statistics derived from image analysis*, Interspeech 2012
- P. Perrot, G. Aversano, R. Blouet, M. Charbit and G. Chollet, *Voice forgery using ALISP: indexation in a client memory*, ICASSP 20015
- D. Matrouf, J.-F. Bonastre and C. Fredouille, *Effect of speech transformation on impostor acceptance*, ICASSP 2006
- T. Kinnunen, Z. Wu, K. A. Lee, F. Sedlak, E. S. Chng and H. Li, *Vulnerability of speaker verification systems against voice conversion spoofing attacks: the case of telephone speech*, ICASSP 2012
- Z. Wu, T. Kinnunen, E. S. Chng, H. Li and E. Ambikairajah, *A study on spoofing attack in state-of-the-art speaker verification: the telephone speech case*, APSIPA ASC 2012
- F. Alegre, A. Amehraye and N. Evans, *A one-class classification approach to generalised speaker verification spoofing countermeasures using local binary patterns*, IEEE BTAS 2013
- Z. Kons and H. Aronowitz, *Voice transformation-based spoofing of text-dependent speaker verification systems*, Interspeech 2013

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

Nicholas Evans, Tomi Kinnunen and Junichi Yamagishi, *Spoofing and countermeasures for automatic speaker verification*, INTERSPEECH 2013

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

Nicholas Evans, Junichi Yamagishi and Tomi Kinnunen, *Spoofing and countermeasures for speaker verification: a need for standard corpora, protocols and metrics*, IEEE Signal Processing Society Newsletter, May 2013

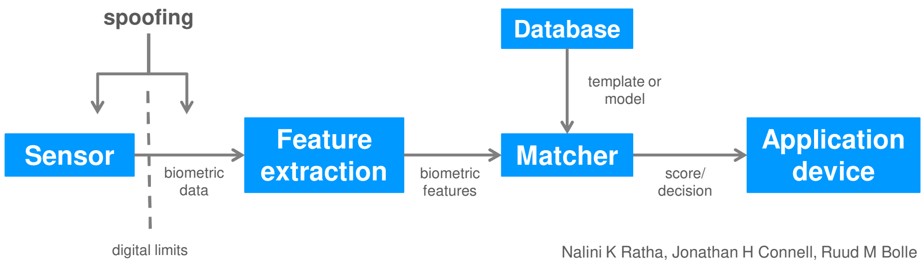
Definitions

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

ISO/IEC 30107-1
http://standards.iso.org/ittf/PubliclyAvailableStandards/c053227_ISO_IEC_30107-1_2016.zip

- *"persons masquerading as others in order to gain illegitimate access to sensitive or protected resources"*

A. Hadid, N. Evans, S. Marcel and J. Fierrez, *Biometrics systems under spoofing attack: an evaluation methodology and lessons learned*, IEEE SPM, 2015

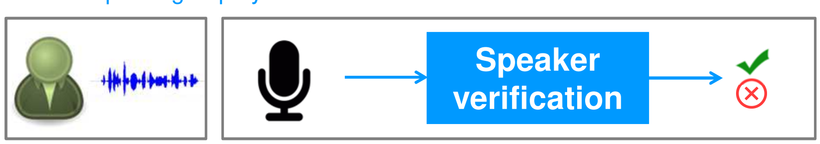


Nalini K Ratha, Jonathan H Connell, Ruud M Bolle
An analysis of minutiae matching strength, Int. Conf. AVBPA, 2001

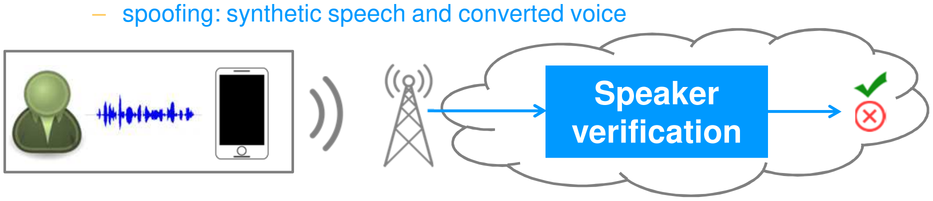
- sensor level: before and after microphone
  - a somewhat contentious issue

Use case scenarios

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



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



Priorities

Spoofing attack	Accessibility	Effectiveness (risk) Text-independent	Effectiveness (risk) Text-dependent	Countermeasure availability
Impersonation	Low	Low/unknown	Low/unknown	Non-existant
Replay	High	Low	Low to high	Low
Speech synthesis	Medium to high	High	High	Medium
Voice conversion	Medium to high	High	High	Medium

ASVspoof 2017

ASVspoof 2015

ASVspoof 2019 ?

ASVspoof 2015

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

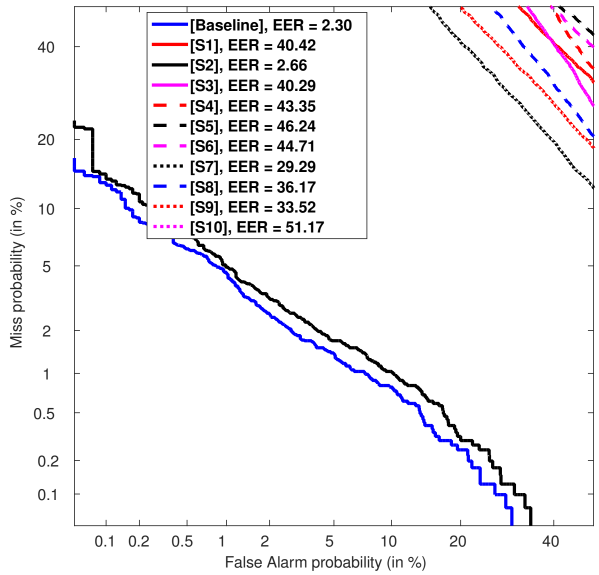
ASVspoof 2015 – spoofing attacks

- **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

	# utterances			Algorithm	Vocoder
	Train	Dev.	Eval.		
Genuine	3750	3497	9404	None	None
S1	2525	9975	18400	VC :Frame-selection	STRAIGHT
S2	2525	9975	18400	VC: Slope-shifting	STRAIGHT
S3	2525	9975	18400	SS: HMM	STRAIGHT
S4	2525	9975	18400	SS: HMM	STRAIGHT
S5	2525	9975	18400	VC: GMM	MLSA
S6	0	0	18400	VC: GMM	STRAIGHT
S7	0	0	18400	VC: GMM	STRAIGHT
S8	0	0	18400	VC: Tensor	STRAIGHT
S9	0	0	18400	VC: KPLS	STRAIGHT
S10	0	0	18400	SS: unit-selection	None

ASVspoof 2015 – vulnerabilities



ivector / PLDA
ASV system

ASVspoof 2015
evaluation set
male subset

ASV

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

Team	Known attacks (S1 - S5)	Unknown attacks (S6 - S10)	Average (all)
DA-IICT	0.408	2.013	1.211
STC	0.008	3.922	1.965
SJTU	0.058	4.998	2.528
NTU	0.003	5.231	2.617
CRIM	0.041	5.347	2.694
F	0.358	6.078	3.218
G	0.405	6.247	3.326
H	0.67	6.041	3.355
I	0.005	7.447	3.726
J	0.025	8.168	4.097
K	0.21	8.883	4.547
L	0.412	13.026	6.719
M	8.528	20.253	14.391
N	7.874	21.262	14.568
O	17.723	19.929	18.826
P	21.206	21.831	21.518

best performance overall & for S10

best performance for S1 – S9

28 teams requested data

16 teams submitted results

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 ceptsral, 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

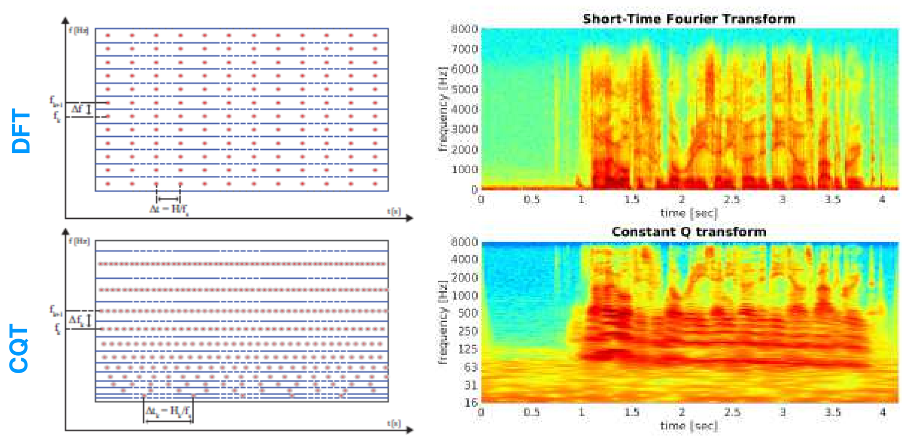
    S. Novoselow, A. Kozlov, G. Lavrentyeva, K. Simonchik and V. Shchemelinin, *STC anti-spoofing systems for the ASVspoof 2015 challenge*, Interspeech 2015
- **SJTU**
  - deep learning, i-vector / PLDA

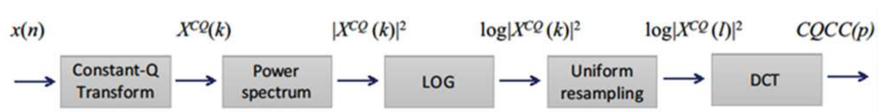
    N. Chen, Y. Qian, H. Dinkel, B. Chen and K. Yu, *Robust deep features for spoofing detection – the SJTU systems for ASVspoof 2015 challenge*, Interspeech 2015
- **NTU**
  - high-resolution phase and magnitude features, MLP

    X. Xiao, X. Tian, S. Du, H. Xu, E. S. Chng and H. Li, *Spoofing speech detection using high dimensional magnitude and phase features: the NTU approach for ASVspoof 2015 challenge*, Interspeech 2015

- Z. Wu, T. Kinnunen, N. Evans, J. Yamagishi, C. Hanilci, M. Sahidullah and A. Sizov, *ASVspoof 2015: the First Automatic Speaker Verification Spoofing and Countermeasures Challenge*
- M. J. Alam, P. Kenny, G. Bhattacharya and T. Stafylakis, *Development of CRIM System for the Automatic Speaker Verification Spoofing and Countermeasures Challenge 2015*
- 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 Countermeasure 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. Sanchez, I. Saratxaga, I. Hernaez, E. Navas and D. Erro, *The AHOLAB RPS SSD Spoofing Challenge 2015 submission*
- 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*
- X. Xiao, X. Tian, S. Du, H. Xu, E. S. Chng and H. Li, *Spoofing Speech Detection Using High Dimensional Magnitude and Phase Features: the NTU Approach for ASVspoof 2015 Challenge*
- S. Novoselov, A. Kozlov, G. Lavrentyeva, K. Simonchik, V. Shchemelinin, *STC Anti-spoofing Systems for the ASVspoof 2015 Challenge*, arXiv:1507.08074, 2015
- S. Weng, S. Chen, L. Yu, X. Wu, W. Cai, Z. Liu and M. Li, *The SYSU System for the Interspeech 2015 Automatic Speaker Verification Spoofing and Countermeasures Challenge*, arXiv:1507.06711, 2015

ASVspoof 2015 – CQCCs





CQCC results for ASVspoof

	Known Attacks						Unknown Attacks						All
System	S1	S2	S3	S4	S5	Avg.	S6	S7	S8	S9	S10	Avg.	Avg.
CFCC-IF	0.101	0.863	0.000	0.000	1.075	0.408	0.846	0.242	0.142	0.346	8.490	2.013	1.211
i-vector	0.004	0.022	0.000	0.000	0.013	0.008	0.019	0.000	0.015	0.004	19.57	3.922	1.965
DNN feat.	0.032	0.109	0.032	0.032	0.086	0.058	0.173	0.049	0.121	0.049	24.601	4.998	2.528
LFCC-DA	0.027	0.408	0.000	0.000	0.114	0.110	0.149	0.011	0.074	0.027	8.185	1.670	0.890
CQCC-A	0.005	0.106	0.000	0.000	0.130	0.048	0.098	0.064	1.033	0.053	1.065	0.462	0.255

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

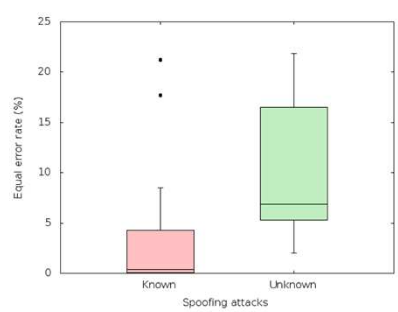
Winner of best paper award
at the Speaker and Language Recognition Workshop (ODYSSEY) 2016

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



Z. Wu, T. Kinnunen, N. Evans, J. Yamagishi, C. Hanilici, M. Sahidullah and A. Sizov,
ASVspoof 2015: the first automatic speaker verification spoofing and countermeasures challenge, Interspeech 2015

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

OCTAVE – EU H2020

Objective Control of Talker Verification

- 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

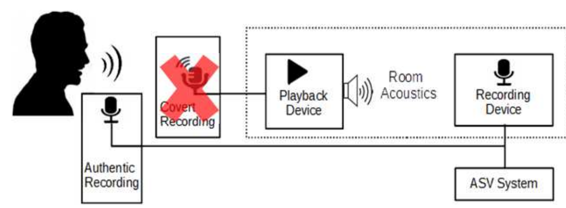
    dynamically increasing database size
- speaker-independent replay detection
  - is this recording authentic or replayed one ?

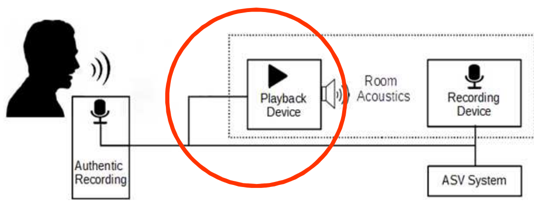
    most general - but can it be done?

Spoofing attacks in automatic speaker verification: analysis and countermeasures 144

Data collection

A realistic scenario





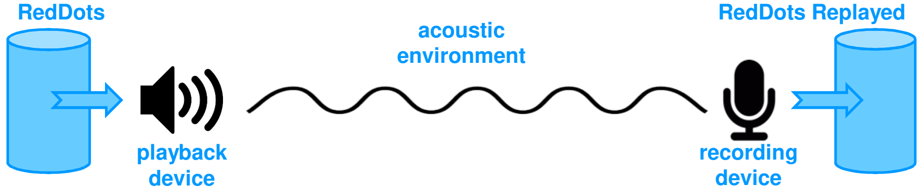
Worst case / simulated scenario

Attacks use a digital copy of a previous ASV access attempt

Allows the use of *existing speech corpora* to create replay attacks

T. Kinnunen, M. Sahidullah, M. Falcone, L. Costantini, R. G. Hautamäki, D. Thomsen, A. Sarkar, Z. H. Tan, H. Delgado, M. Todisco, N. Evans, V. Hautamäki and K. A. Lee, *RedDots Replayed: A New Replay Spoofing Attack Corpus for Text-dependent Speaker Verification Research*, ICASSP 2017

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/

K. A. Lee, A. Larcher, G. Wang, P. Keny, N. Brümmer, D. A. van Leeuwen, H. Aronowitz, M. Kockmann, C. Vaquero, B. Ma, H. Li, T. Stafylakis, M. J. Alam, A. Swart and J. Perez, *The RedDots data collection for speaker recognition*, Interspeech 2015

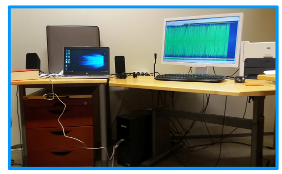
Example replay configurations

Smartphone → Smartphone

Headphones
→ PC mic

High-quality loudspeaker
→ smartphone, anechoic room

High-quality loudpspeaker
→ high-quality mic



Laptop line-out
→ PC line-in using a cable

Results

ASV results (Equal error rate, EER %)

Type of impostor	GMM-UBM	i-vector (cosine)	i-vector (PLDA)
Zero effort	2.50	6.64	5.23
Replay	23.18	26.63	24.85

speakers: 62
data: 3854 genuine,
52944 spoofed

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

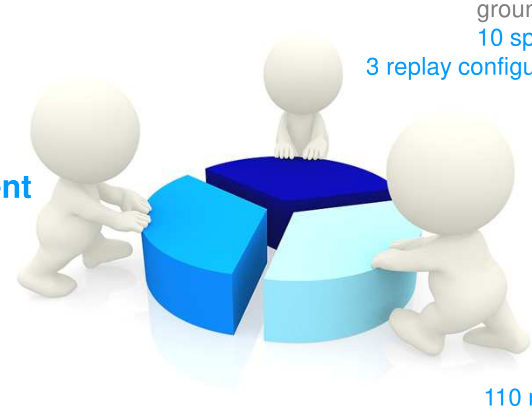
Front-end feature	Controlled	Variable	All
LFCC 20-da	5.88	4.43	5.11
CQCC 20-a	2.77	3.50	3.27

T. Kinnunen, M. Sahidullah, M. Falcone, L. Costantini, R. G. Hautamäki, D. Thomsen, A. Sarkar, Z. H. Tan, H. Delgado, M. Todisco, N. Evans, V. Hautamäki and K. A. Lee, *RedDots Replayed: A New Replay Spoofing Attack Corpus for Text-dependent Speaker Verification Research*, ICASSP 2017

ASVspoof 2017

ASVspoof 2017

- **physical access and replay spoofing attacks**

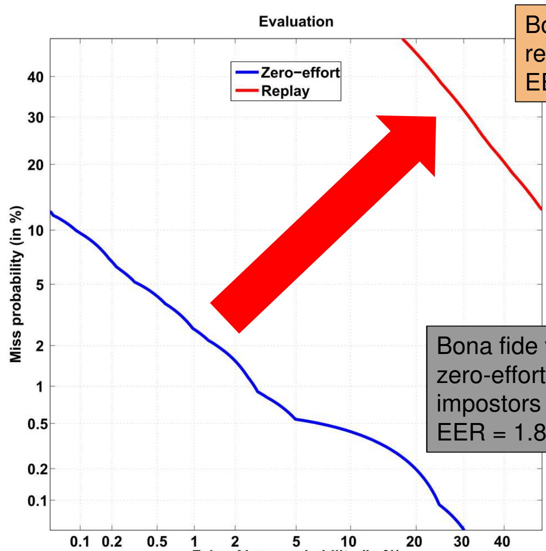


ASVspoof 2017 – dimensions

Subset	# Speakers	# Replay sessions	# Replay configs	# utterances	
				Bona fide	Replay
Training	10	6	3	1508	1508
Development	8	10	10	760	950
Evaluation	24	161	110	1298	12008
Total	42	177	123	3566	14466

Largest, most diverse replay dataset

ASVspoof 2017 – vulnerabilities



Bona fide vs. replay spoofs EER = 31.5 %

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

GMM-UBM ASV system

ASVspoof 2017 evaluation set

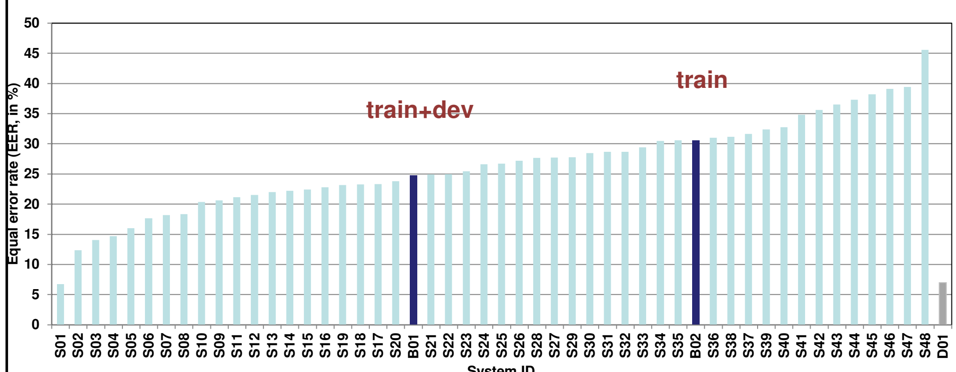
ASV

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

Sxx: Regular submission
Bxx: Baseline system
Dxx: Late submission



- 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

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

CQCC baseline features

DNN-based classifier
Other classifier

T: training
T+D: training + development

- T. Kinnunen, M. Sahidullah, H. Delgado, M. Todisco, N. Evans, J. Yamagishi and K. A. Lee, *The ASVspoof 2017 Challenge: Assessing the Limits of Replay Spoofing Attack Detection*
- R. Font, J. M. Espín and M. J. Cano, *Experimental Analysis of Features for Replay Attack Detection — Results on the ASVspoof 2017 Challenge*
- H. A. Patil, M. R. Kamble, T. B. Patel and M. H. Soni , *Novel Variable Length Teager Energy Separation Based Instantaneous Frequency Features for Replay Detection*
- W. Cai, D. Cai, W. Liu, G. Li and M. Li , *Countermeasures for Automatic Speaker Verification Replay Spoofing Attack : On Data Augmentation, Feature Representation, Classification and Fusion*
- S. Jelil, R. K. Das, S. R. M. Prasanna and R. Sinha, *Spoof Detection Using Source, Instantaneous Frequency and Cepstral Features*
- M. Witkowski, S. Kacprzak, P. Żelasko, K. Kowalczyk and J. Gałka, *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*
- G. Lavrentyeva, S. Novoselov, E. Malykh, A. Kozlov, O. Kudashev and V. Shchemelinin, Audio Replay Attack Detection with Deep Learning Frameworks
- 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*
- L. Li, Y. Chen, D. Wang and T. F. Zheng, *A Study on Replay Attack and Anti-Spoofing for Automatic Speaker Verification*
- P. Nagarsheth, E. Khoury, K. Patil and M. Garland, *Replay Attack Detection Using DNN for Channel Discrimination*
- Z. Chen, Z. Xie, W. Zhang and X. Xu, *ResNet and Model Fusion for Automatic Spoofing Detection*
- K. N. R. K. R. Alluri, S. Achanta, S. R. Kadiri, S. V. Gangashetty and A. K. Vuppala, *SFF Anti-Spoofer: IIIT-H Submission for Automatic Speaker Verification Spoofing and Countermeasures Challenge 2017*

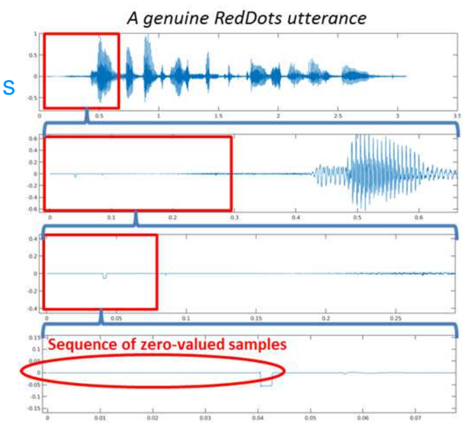
ASVspoof 2017 – summary

- 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
  - genearlisation 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 characteristisations
  - 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

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

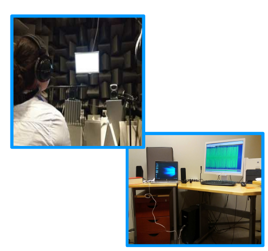
Generic IDs split into specific IDS

Example: *Home (generic) → Home 01, Home 02, …*

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



ASVspoof 2017 – playback devices

ID	Playback device
P01	All-in-one PC speakers
P02	Creative A60 speakers
P03	Genelec 8020C studio monitor
P04	Genelec 8020C studio monitor (2 speakers)
P05	Beyerdynamic DT 770 PRO headphones
P06	Dell laptop internal speakers
P07	Dynaudio BM5A speaker
P08	HP Laptop internal speakers
P09	VIFA M10MD-39-08 speaker
P10	ACER netbook internal speakers
P11	BQ Aquaris M5 smartphone
P12	Logitech low quality speakers
P13	Desktop PC line output
P14	Labtec LCS-1050 speakers
P15	Edirol MA-15D studio monitor
P16	Lenovo Ideatab S6000-H tablet
P17	Logitech S120 multimedia speakers
P18	MacBook pro internal speakers
P19	Altec lansing Orbit USB iML227 portable speaker
P20	Samsung GT-I9100 smartphone
P21	Samsung GT-P6200 tablet
P22	Behringer Truth B2030A studio monitor
P23	Focusrite Scarlett 2i2 audio interface line output
P24	Focusrite Scarlett 2i4 audio interface line output
P25	Genelec 6010A studio monitor
P26	AKG K242HD Headset

Devices with small loudspeakers: laptops / smartphones/ tablets

Larger-size, consumer loudspeakers (desktop speakers)

Studio-quality loudspeakers / headphones / analog outputs

ASVspoof 2017 – recording devices

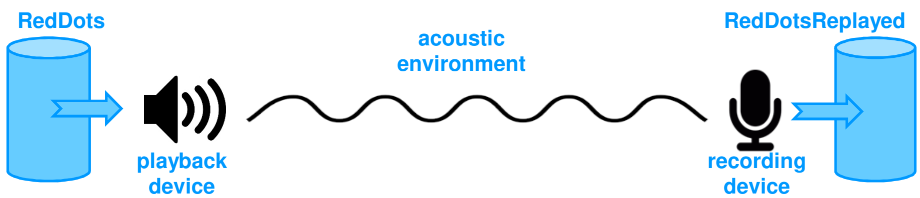
ID	Recording device
R01	Zoom H6 handy recorder
R02	BQ Aquaris M5 smartphone
R03	Low-quality headset
R04	Nokia Lumia 635 smartphone
R05	Røde NT2 microphone
R06	Røde smartLav+ microphone
R07	Samsung Galaxy S7 smartphone
R08	Desktop PC microphone input
R09	Zoom H6 recorder with Behringer ECM8000 mic.
R10	Zoom H6 recorder with MSH-6 microphone
R11	Zoom H6 recorder. with XY microphone
R12	iPhone 5c smartphone
R13	iPhone 7 plus smartphone
R14	iPhone 4 smartphone
R15	Logitech C920 webcam
R16	miniDSP UMIK-1 microphone
R17	Samsung Galaxy Trend 2 smartphone
R18	Samsung GT-I9100 smartphone
R19	Samsung GT-P6200 tablet
R20	Samsung Trend 2 smartphone
R21	AKG C3000 microphone
R22	SE electronic 2200a microphone
R23	Focusrite Scarlett 2i2 interface line input
R24	Focusrite Scarlett 2i4 interface line input
R25	Zoom HD1 handy recorder

Devices with small microphones: smartphones / tablets

Headset / webcam microphones

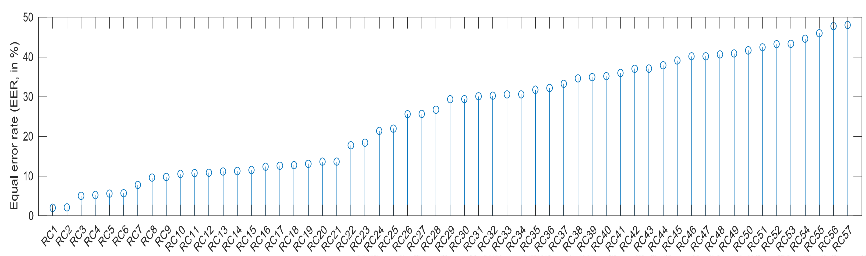
Studio-quality microphones / hand-held recorders / analog inputs

ASVspoof 2017 – replay configurations



Subset	# Speakers	# Replay sessions	# Replay configs	# utterances Bona fide	# utterances Replay
Training	10	6	3	1507	1507
Development	8	10	10	760	950
Evaluation	24	161	~~110~~ 57	1298	12008
Total	42	177	~~123~~ 61	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

ID	E	P	R	#seg.	ID	E	P	R	#seg.	ID	E	P	R	#seg.
RC01	18	18	12	55	RC20	7	21	14	275	RC39	22	3	4	183
RC02	8	20	19	67	RC21	18	11	12	22	RC40	16	7	6	116
RC03	8	21	25	122	RC22	12	16	11	183	RC41	13	14	9	182
RC04	8	20	25	102	RC23	20	10	15	1138	RC42	11	26	16	153
RC05	8	20	14	114	RC24	17	12	17	179	RC43	6	9	6	84
RC06	8	21	14	108	RC25	9	18	12	42	RC44	13	14	10	179
RC07	8	21	18	120	RC26	17	12	10	184	RC45	10	25	16	346
RC08	2	21	25	98	RC27	6	9	7	96	RC46	18	22	21	181
RC09	7	20	25	244	RC28	21	3	1	240	RC47	23	15	13	342
RC10	2	20	25	82	RC29	18	5	3	454	RC48	19	22	22	1200
RC11	7	20	19	272	RC30	15	19	20	74	RC49	14	3	17	180
RC12	2	20	19	75	RC31	16	7	7	145	RC50	1	15	13	748
RC13	3	8	20	113	RC32	17	12	9	180	RC51	16	7	5	169
RC14	7	21	18	279	RC33	13	14	4	183	RC52	10	26	16	181
RC15	2	21	18	150	RC34	17	12	4	181	RC53	14	3	4	181
RC16	2	21	14	116	RC35	13	14	17	178	RC54	6	9	5	105
RC17	7	21	25	266	RC36	22	4	17	181	RC55	26	24	24	178
RC18	7	20	14	265	RC37	15	19	4	48	RC56	25	13	8	182
RC19	2	20	14	120	RC38	12	17	11	184	RC57	24	23	23	183

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

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

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

G. Lavrentyeva, S. Novoselov, E. Malykh, A. Kozlov, O. Kudashev, V. Shchemelinin, *Audio Replay Attack Detection with Deep Learning Frameworks, INTERSPEECH*, 2017

R. Font, J.M. Espín, M.J. Cano, *Experimental Analysis of Features for Replay Attack Detection — Results on the ASVspoof 2017 Challenge,.INTERSPEECH*, 2017

ASVspoof 2017 – performance

Spoofing detection performance (EER,%)

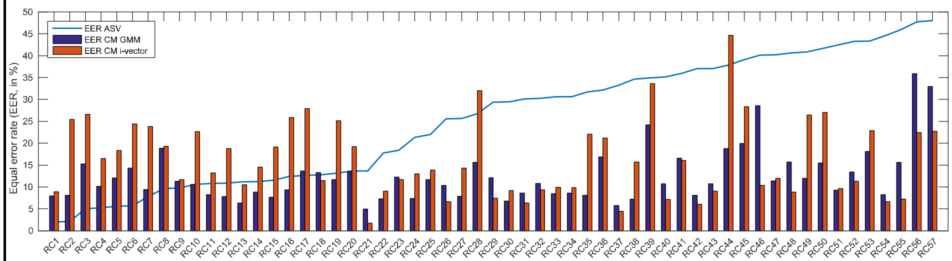
	training on	T	D	T	D	T+D	T	D	T	D	T+D
	testing on	D	T	E			D	T	E		
	Feat config.	no normalisation					CMVN				
GMM	19-SDA	11.69	1.36	30.79	25.33	23.97	13.31	8.49	19.74	16.89	15.33
	19E-SDA	10.37	1.37	34.95	26.3	29.31	**9.06**	**5.64**	**13.74**	**14.77**	**12.24**
i-vector	19-SDA	**4.43**	**1.23**	**17.82**	**18.81**	**18.60**	11.61	8.74	16.61	15.08	15.63
	19E-SDA	5.11	1.54	21.47	16.25	21.10	10.52	7.27	14.76	14.37	12.93

T Train
D Development
E Evaluation
CMVN Cesptral mean and variance normalisation
19E With energy coefficient
SDA static + delta + acceleration coefficients

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

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

ASVspoof 2017 – RC analysis



- **no correlation between ASV degradation and replay detection performance**

	Low	Medium	High
E: acoustic environment	16.68	18.73	21.86
P: playback device	16.64	16.44	18.37
R: recording device	10.80	15.69	17.77

Acoustic environment seems to have a greater impact on performance

- **inconsistent performance for GMM and i-vector CMs**

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

the future

ASVspoof 2019

ASVspoof 2019

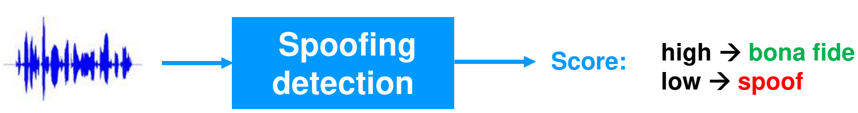
- logical access AND physical access scenarios
  - state-of-the-art synthetic speech, converted voice and replay
- controlled setup
- based upon VCTK
  - same as ASVspoof 2015
- protocol enhancements
- impact upon ASV
  - **metrics and integration**

Spoofing attacks in automatic speaker verification: analysis and countermeasures 170

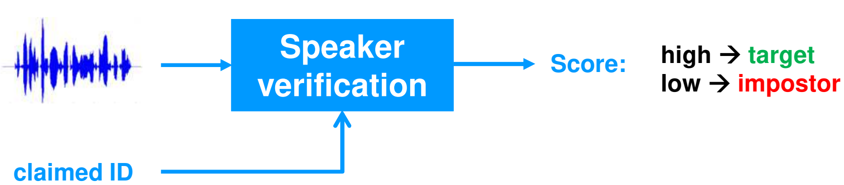
Metrics and integration

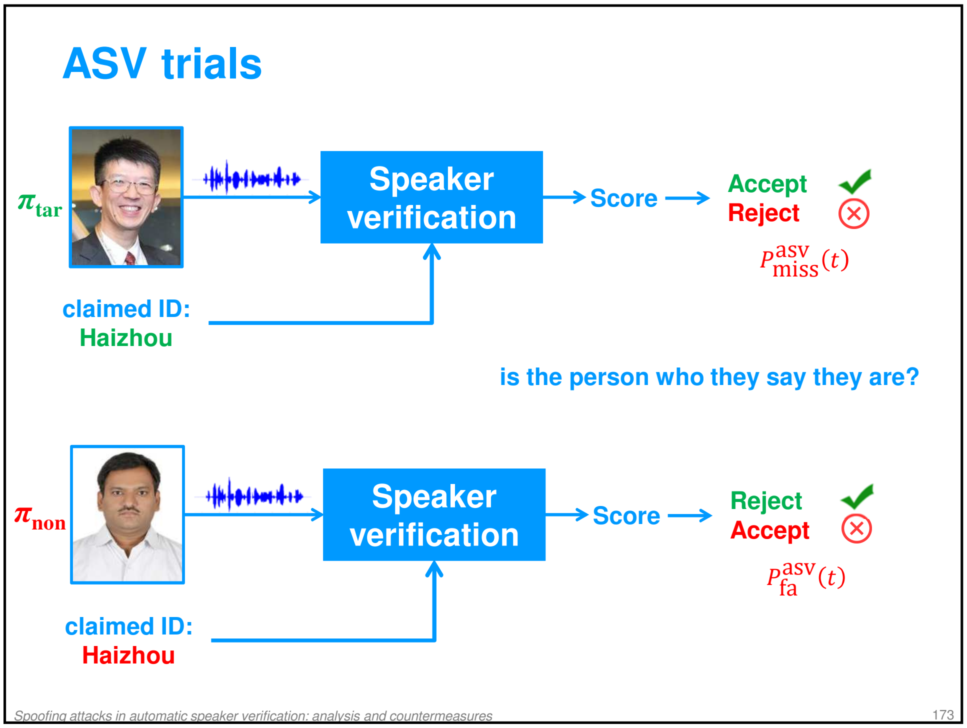
- limitations of the previous work
  - independent spoofing countermeasures
  - does not reflect impact upon ASV
- vision for the future
  - reflect integrated systems
  - backward compatibility with standalone assessment
  - allows the specification of **application** costs and priors
  - facilitates unified comparison of
    - ASV without countermeasure
    - ASV with perfect countermeasure
    - perfect ASV system with countermeasure
  - metric that is easy to understand and use

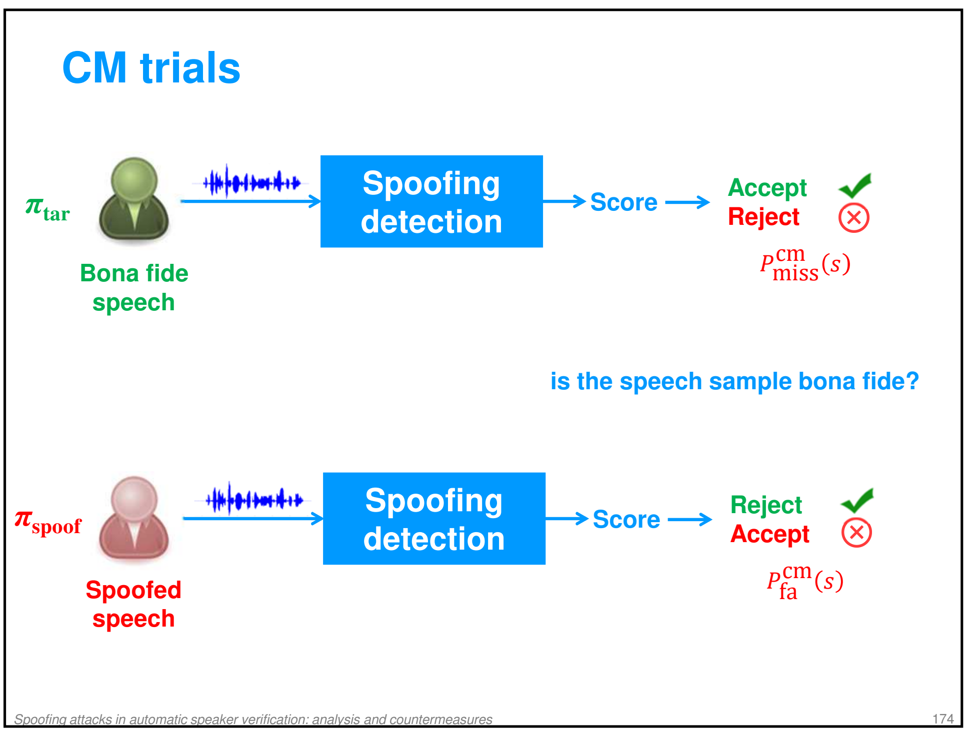
Integration



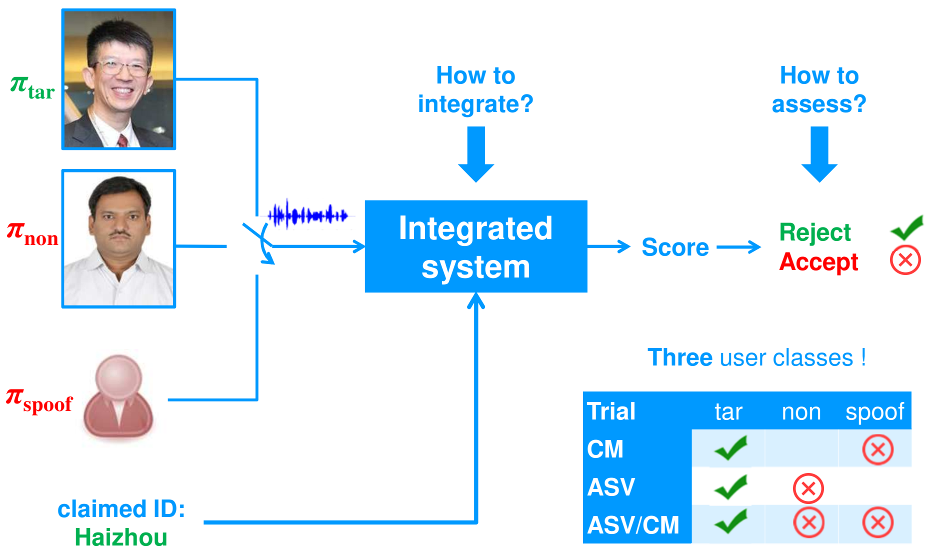
two systems – different objectives – how many user classes?




ASV trials
π_{tar}
Speaker verification
Score
Accept
Reject
$P_{\text{miss}}^{\text{asv}}(t)$
claimed ID:
Haizhou
is the person who they say they are?
π_{non}
Speaker verification
Score
Reject
Accept
$P_{\text{fa}}^{\text{asv}}(t)$
claimed ID:
Haizhou
Spoofing attacks in automatic speaker verification: analysis and countermeasures
173

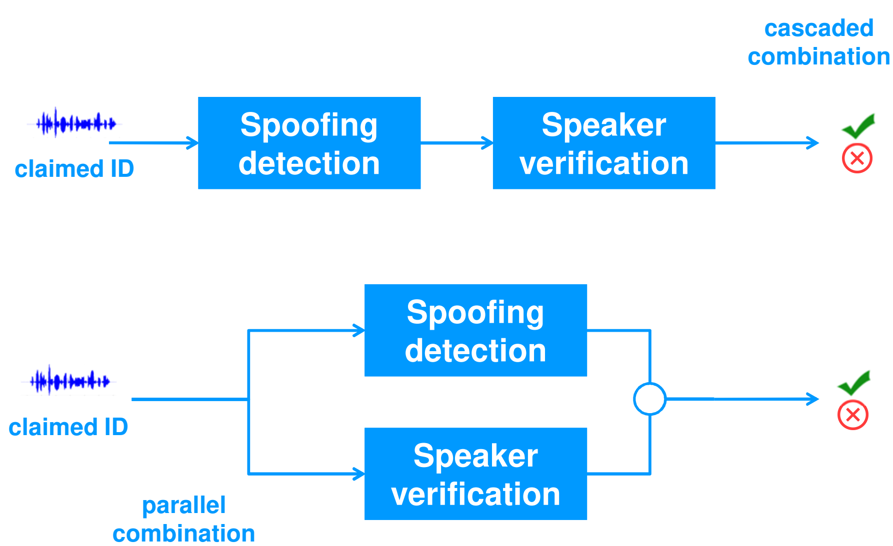

CM trials
π_{tar}
Bona fide speech
Spoofing detection
Score
Accept
Reject
$P_{\text{miss}}^{\text{cm}}(s)$
is the speech sample bona fide?
π_{spoof}
Spoofed speech
Spoofing detection
Score
Reject
Accept
$P_{\text{fa}}^{\text{cm}}(s)$
Spoofing attacks in automatic speaker verification: analysis and countermeasures
174

Integrated system trials

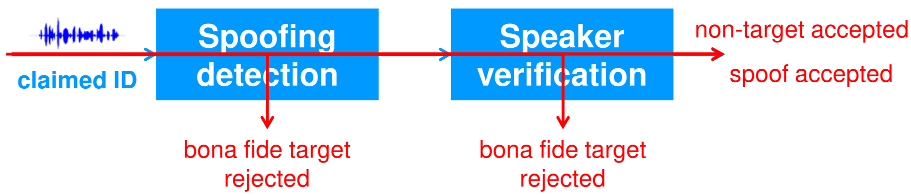


Trial	tar	non	spoof
CM	✓		⊗
ASV	✓	⊗	
ASV/CM	✓	⊗	⊗

Architectures



Integrated system errors



Four possible errors

ASV system rejects target

$$P_{\mathrm{a}}(s,t) = (1 - P^{\mathrm{cm}}_{\mathrm{miss}}(s)) \times P^{\mathrm{asv}}_{\mathrm{miss}}(t)$$

CM falsely accepts spoof

$$P_{\mathrm{c}}(s,t) = P^{\mathrm{cm}}_{\mathrm{fa}}(s) \times (1 - P^{\mathrm{asv}}_{\mathrm{miss}}(t))$$

ASV system accepts non-target

$$P_{\mathrm{b}}(s,t) = (1 - P^{\mathrm{cm}}_{\mathrm{miss}}(s)) \times P^{\mathrm{asv}}_{\mathrm{fa}}(t)$$

CM rejects target

$$P_{\mathrm{d}}(s) = P^{\mathrm{cm}}_{\mathrm{miss}}(s)$$

t-DCF

The traditional NIST-defined detection cost function at ASV threshold *t*

$$\mathrm{DCF}(t) = C^{\mathrm{asv}}_{\mathrm{miss}}\, \pi_{\mathrm{tar}}\, P^{\mathrm{asv}}_{\mathrm{miss}}(t) + C^{\mathrm{asv}}_{\mathrm{fa}}\, \pi_{\mathrm{non}}\, P^{\mathrm{asv}}_{\mathrm{fa}}(t)$$

$$\pi_{\mathrm{tar}} + \pi_{\mathrm{non}} = 1$$

The tandem detection cost function at ASV threshold *t* and CM threshold *s*

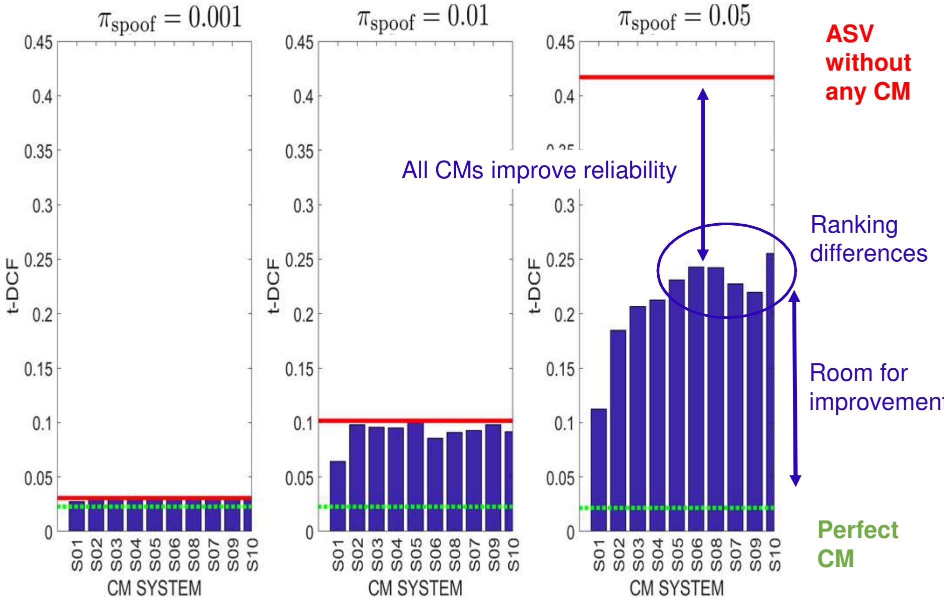
$$\mathrm{t{-}DCF}(s,t) = C^{\mathrm{asv}}_{\mathrm{miss}}\, \pi_{\mathrm{tar}}\, P_{\mathrm{a}}\,(s,t) + C^{\mathrm{asv}}_{\mathrm{fa}}\, \pi_{\mathrm{non}}\, P_{b}\,(s,t)$$

$$+ C^{\mathrm{cm}}_{\mathrm{fa}}\, \pi_{\mathrm{spoof}}\, P_{c}\,(s,t) + C^{\mathrm{cm}}_{\mathrm{miss}}\, \pi_{\mathrm{tar}}\, P_{\mathrm{d}}\,(s)$$

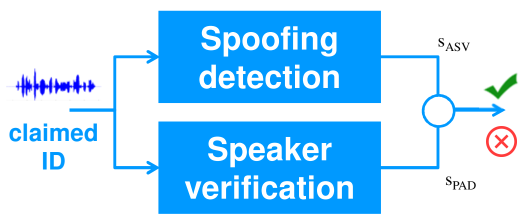
$$\pi_{\mathrm{tar}} + \pi_{\mathrm{non}} + \pi_{\mathrm{spoof}} = 1$$

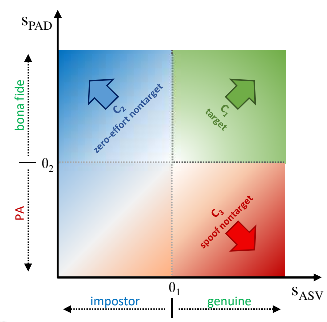
T. Kinnunen, K. Aik Lee, H. Delgado, N. Evans, M. Todisco, M. Sahidullah, J. Yamagishi and D. A. Reynolds, *t-DCF: a detection cost function for the tandem assessment of spoofing countermeasures and automatic speaker verification*, ODYSSEY 2018

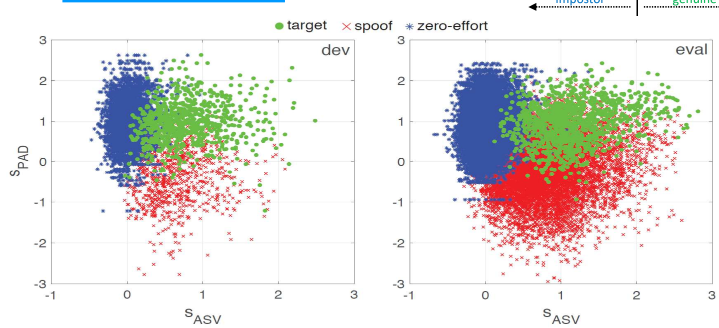
t-DCF: ASVspoof 2017



Alternative integration







M. Todisco, H. Delgado, K. A. Lee, Md. Sahidullah, N. Evans, T. Kinnunen and J. Yamagishi, *Integrated Presentation Attack Detection and Automatic Speaker Verification: Common Features and Gaussian Back-end Fusion*, Interspeech 2018

ASVspoof 2019 – tentative schedule

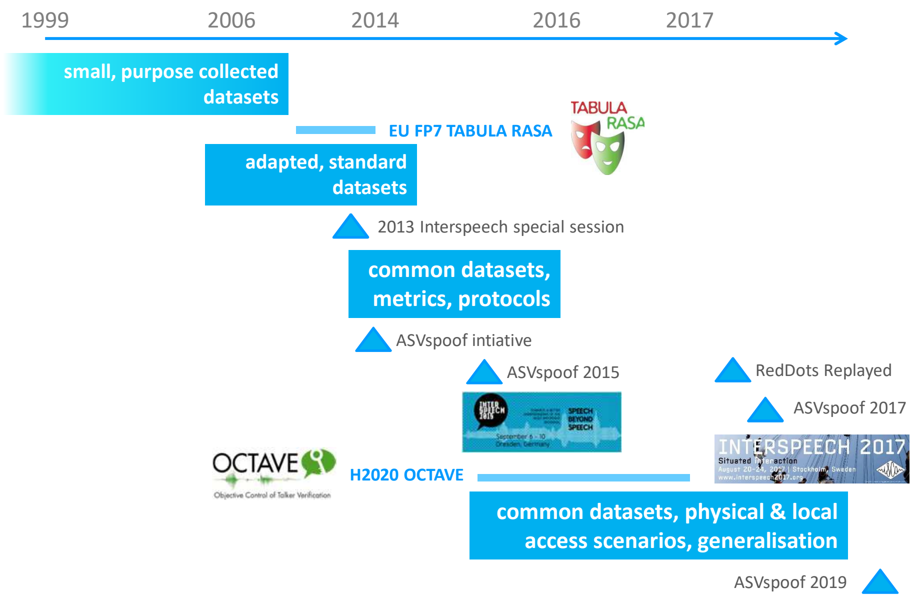
new slide

Development data	late November to early December
Evaluation data	early-to-mid February
Score submission	+1 week
Results	end of February
Interspeech deadline	**29th March**

Stay tuned at:
http://www.asvspoof.org

Wrap up

Overview



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 !!

Spoofing attacks in automatic speaker verification: analysis and countermeasures 184

New publications

- INTERSPEECH 2018 – oral session Monday 3rd 15.00, Hall 2: 15:00
  - M.Todisco, H. Delgado, K. A. Lee, M. Sahidullah, N. Evans, T. Kinnunen and J. Yamagishi, *Integrated Presentation Attack Detection and Automatic Speaker Verification: Common Features and Gaussian Back-end Fusion*
- 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. Jelil, S. Kalita, S. R. M. Prasanna and R. Sinha, *Exploration of Compressed ILPR Features for Replay Attack Detection*
  - T. Gunendradasan, B. Wickramasinghe, N. P. Le, E. Ambikairajah and J. Epps, *Detection of Replay-Spoofing Attacks Using Frequency Modulation Features*
  - M. Kamble, H. Tak and H. Patil, *Effectiveness of Speech Demodulation-Based Features for Replay Detectio*
  - M. Kamble and H. Patil, *Novel Variable Length Energy Separation Algorithm Using Instantaneous Amplitude Features for Replay Detection*
  - J. Yang, C. You and Q. He, *Feature with Complementarity of Statistics and Principal Information for Spoofing Detection*
  - D. Li, L. Wang, J. Dang, M. Liu, Z. Oo, S. Nakagawa, Haotian 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*
  - H. Sailor, M. Kamble and H. Patil, *Auditory Filterbank Learning for Temporal Modulation Features in Replay Spoof Speech Detection*
  - K. Sriskandaraja, V. Sethu and E. Ambikairajah, *Deep Siamese Architecture Based Replay Detection for Secure Voice Biometric*
  - A. G. Alanís, A. M. Peinado, J. A. Gonzalez and A. Gomez, *A Deep Identity Representation for Noise Robust Spoofing Detection*
  - F. Tom, M. Jain and P. Dey, *End-To-End Audio Replay Attack Detection Using Deep Convolutional Networks with Attention*
  - Saranya M. S. and H. Murthy, *Decision-level Feature Switching as a Paradigm for Replay Attack Detection*
  - G. Suthokumar, V. Sethu, C. Wijenayake 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
  - P. Tapkir and H. Patil, *Novel Empirical Mode Decomposition Cepstral Features for Replay Spoof Detection*
  - 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
  - M. Sahidullah, H. Delgado, M. Todisco, T. Kinnunen, N. Evans, J. Yamagishi and K. A. Lee, *Introduction to Voice Presentation Attack Detection and Recent Advances*

Speaker recognition software

- **ALIZE 3.0**
  http://www1.i2r.a-star.edu.sg/~alarcher/Softwares.html
- **SPEAR Toolkit (based on BOB)**
  https://pypi.python.org/pypi/bob.spear/1.9.0, http://idiap.github.io/bob/
- **MSRidentity Toolbox**
  https://www.microsoft.com/en-us/research/wp.../MSR-Identity-Toolbox-v1_1.pdf
- **Kaldi**
  https://github.com/kaldi-asr/kaldi
- **Sidekit**
  http://www-lium.univ-lemans.fr/sidekit/

Databases

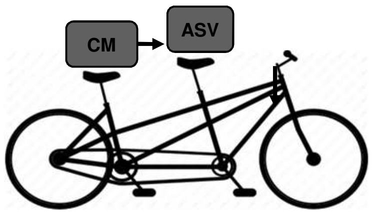
- **ASVspoof**
  http://www.asvspoof.org
- **NIST speaker recognition evaluation corpora available from Linguistic Data Consortium**
  https://www.ldc.upenn.edu/
- **RSR2015** [Larcher et al, Interspeech '12]
  https://www.etpl.sg/innovation-offerings/ready-to-sign-licenses/rsr2015-overview-n-specifications
- **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.asvspoof.org/data2017/tDCF_v0.1.zip



Thank you for listening

Haizhou Li

NUS National University of Singapore

Hemant A. Patil

DA-IICT

Nicholas Evans

EURECOM Sophia Antipolis