Audio Security & Privacy

Andreas Nautsch
EURECOM
Outline

- Security: ASVspoof challenges
- Privacy: speech meets legal & crypto experts
- ISCA SIG: Security & Privacy in Speech Communication
  // ISCA: Int’l Speech Communication Association
Security in voice biometrics is becoming a necessity
Voice spoofing & biometric presentation attacks

speech synthesis

voice conversion

replay

impersonation

Sneakers (1992)  Universal Pictures

office  meeting  anechoic

mobile phone  HQ loudspeaker  HQ loudspeaker

Andreas here... verify my voice!

mimicry by a human being

greatest threats!
ASVspoof 2015

16 organizations participated

ASVspoof 2017

49 organizations participated
ASVspoof 2019

154 teams participated

154 teams participated

154 teams participated

PA spoofing attack

LA spoofing attack

Sensor (mic)

feature extraction

countermeasure

live human being

score

spoofed sample

speech features

speech data

speech data

speech data

speech data

speech data

speech data

required from participants

decision

set out by organisers

ASV scores

CM scores

countermeasure

ASV

ASV scores
ASVspoof 2019 — Database

● based on VCTK corpus [1]
  ○ omni-directional head-mounted microphone (DPA 4035)
  ○ 96kHz sampling frequency @ 24 bits
  ○ hemi-anechoic chamber of the University of Edinburgh

● common partitions for LA and PA
  ○ 107 English speakers
  ○ speakers for eval, dev and training set
  ○ ASV enrollment

ASVspoof 2019 — Logical access attacks

ASV only zero-effort impostors → EER = 2.48%

<table>
<thead>
<tr>
<th>ASV</th>
<th>44.66</th>
<th>59.68</th>
<th>40.39</th>
<th>8.38</th>
<th>57.73</th>
<th>59.64</th>
<th>46.18</th>
<th>46.78</th>
<th>64.01</th>
<th>58.85</th>
<th>64.52</th>
<th>3.92</th>
<th>7.35</th>
<th>14.58</th>
</tr>
</thead>
<tbody>
<tr>
<td>EER</td>
<td>7.83</td>
<td>0.02</td>
<td>0.09</td>
<td>0.06</td>
<td>12.21</td>
<td>0.59</td>
<td>3.75</td>
<td>12.41</td>
<td>2.88</td>
<td>3.22</td>
<td>0.02</td>
<td>15.93</td>
<td>5.59</td>
<td>0.06</td>
</tr>
</tbody>
</table>

13 attacks breakdown
ASVspoof 2019 — Physical access attacks

ASV only zero-effort impostors → EER = 6.47%

9 attacks breakdown

<table>
<thead>
<tr>
<th>ASV</th>
<th>40.48</th>
<th>45.03</th>
<th>43.98</th>
<th>41.01</th>
<th>40.56</th>
<th>39.96</th>
<th>37.85</th>
<th>38.83</th>
<th>38.16</th>
<th>36.42</th>
</tr>
</thead>
<tbody>
<tr>
<td>CM</td>
<td>8.09</td>
<td>17.01</td>
<td>5.64</td>
<td>2.21</td>
<td>14.32</td>
<td>4.40</td>
<td>1.67</td>
<td>12.99</td>
<td>4.28</td>
<td>1.96</td>
</tr>
</tbody>
</table>

replay device quality
- high
- medium
- low

attacker-to-talker distance
- close
- medium
- far

pooled  AA  AB  AC  BA  BB  BC  CA  CB  CC
ASVspoof 2019 — Physical access attacks

27 environments breakdown

reverberation noise
- low
- medium
- high

room size
- small
- medium
- large

talker-to-ASV distance
- close
- med
- far

min t-DCF

primary T28
B01
B02
single T28
ASVspoof 2019 — “the hidden track of the album”
ASVspoof 2019 — Organisers

Junichi Yamagishi
NII, Japan
Univ. of Edinburgh, UK

Massimiliano Todisco
EURECOM, France

Md Sahidullah
Inria, France

Héctor Delgado
EURECOM, France
Nuance, Spain

Nicholas Evans
EURECOM, France

Xin Wang
NII, Japan

Ville Vestman
UEF, Finland

Kong Aik Lee
NEC, Japan

Tomi H. Kinnunen
UEF, Finland

Andreas Nautsch
EURECOM, France

Md Sahidullah
Inria, France
— Privacy —

speech meets legal & crypto experts
Preserving privacy in speaker and speech characterisation ★

Andreas Nautsch, Abelino Jiménez, Amos Treiber, Jascha Kolberg, Catherine Jasserand, Els Kindt, Héctor Delgado, Massimiliano Todisco, Mohamed Amine Hmani, Aymen Mtibaa, Mohammed Ahmed, Abdelraheem, Alberto Abad, Francisco Teixeira, Driss Matrouf, Marta Gomez-Barrero, Dijana Petrovska, Delacéritaz, Gérard Chollet, Nicholas Evans, Thomas Schneider, Jean-François Bonastre, Bhiksha Raj, Isabel Trancoso, Christoph Busch

Speaker recognition          Study of the Law          Biometrics
Speech communication        Cryptography
Why is speech data sensitive?

“Speech is a medium in communication to impart or exchange information.”

- Characteristics
  - Behavioural
  - Physiological
  - What we say

- Data types
  - Audio
  - Text
  - Video
  - Brainwaves
  - ...

https://www.eslfast.com/robot/audio/dailylife/dailylife1901.mp3
Privacy & speech data, a legal perspective  I/III

- There is not a single or universal legal definition of “privacy” (!)

- Warren and Brandeis (US, 1890): “the right to be let alone”

- US: 4 types of privacy
  - Informational privacy ⇒ data privacy
  - Physical privacy
  - Decisional privacy
  - Proprietary privacy

- EU: “broad term not susceptible to exhaustive definition”
  - Art. 8 European Convention on Human Rights
  - Art. 7 Catalogue of Fundamental Rights and Freedoms
Privacy & speech data, a legal perspective  II/III

● Law in the US
  ○ Acts/provisions in California, Illinois, Texas & Washington
  ○ Illinois & Texas: restrictive definition ‘biometric identifier’
  ○ Washington: definition by examples, e.g., ‘voiceprints’
  ○ 2020: California Consumer Privacy Act
     ‘identifiers’ can be extracted from ‘biometric information’

● Law in the EU
  ○ European Convention of Human Rights
  ○ Catalogue of Fundamental Rights and Freedoms
  ○ GDPR 2016/679
  ○ Police Directive 2016/680
  ○ Payment services directive (PSD 2)
  ○ ePrivacy regulation (under implementation)
Privacy & speech data, a legal perspective III/III

- European perspective

- Biometric data is not the sole "sensitive data"
  - Racial/ethnic origin
  - Political opinions
  - Religious/philosophical beliefs
  - Health data

- Data Protection Impact Assessment (DPIA)
  - Required for processing ‘on a large scale’
  - Obligation of ‘controllers’ and ‘processors’
  - Technical & organisational measures ensuring
  - Evaluating the effectiveness of security measures: confidentiality, integrity, availability & resilience
Privacy by Design & Privacy by Default

- EU GDPR
  - Technical/organisational measures beyond security measures
  - Factors:
    - State-of-the-art (standards, research, ...)
    - Cost of implementation
    - Nature, scope, context & purpose of processing
    - Risks to individuals’ rights
  - Limitation of data collection to what is ‘strictly necessary’
  - ‘By design’: policy principle
  - ‘By design and by default’: legal obligation
Resources provided by the EDPS

● EDPS: European Data Protection Supervisor
  ○ Handbook on European data protection law
  ○ EDPS TechDispatch ⇒ TechDispatch #1: Smart Speakers and Virtual Assistants
  ○ EDPS Website Evidence Collector
  ○ Introduction to the hash function as a personal data pseudonymisation technique
  ○ EDPS Preliminary Opinion on Privacy by Design
  ○ EDPB Guidelines 4/2019 on Article 25 Data Protection by Design and by Default
  ○ EDPS IPEN workshops

● [Slides] https://www.spsc-sig.org/2020-01-29-speech-legal-workshop
  Talk of Thomas Zerdick, Head of Unit “IT-policy” @ EDPS

● “Data protection” = using safeguards for sensitive information
Privacy & speech data; cybersecurity  I/III

● So ... which “safeguards” do we have?

  in other words, which cryptographic approaches are proposed?
  ○ HE: homomorphic encryption ← covered
  ○ STPC: secure two-party computation ← in this talk
  ○ DP: differential privacy
  ○ FL: federated learning
  ○ Intel SGX: hardware-assisted security

● How to check, we did well?

● Spoiler: always have a crypto expert around — plenty of space for mistakes
Privacy & speech data; cybersecurity II/III

- Odyssey 2018: HE for speaker recognition
  \[ Enc_{pk}(x) \oplus Enc_{pk}(y) = Enc_{pk}(x+y) \]
  "compute all-at-once" approach
  Slow computation
  Low communication

- Interspeech 2019: HE & STPC

- Speech Communication 2020: STPC
  "compute bit-by-bit" approach
  Fast computation
  High communication

- Note: related work by Rahulamathavan et al. (CyberSA’18 & TASLP’19)
  But: found to be highly insecure by Schneider & Treiber (TPDS’20)
Privacy & speech data; cybersecurity III/III

Fundamentals

Assumptions
Cryptography & secure computation

Method

Proofs & implementations
Theoretic: very strong
Empirical: weaker

Goals

How to prepare functions and data, such that the correct outcome can be derived from a computation which operates on protected data?

⇒ Formal definitions

Computations

Infrastructure

Communication

Zero knowledge “zero evidence”

Computational indistinguishability

Cryptographic hardness
Privacy & speech data; cybersecurity — easy, right?

Fundamentals

Assumptions
- Cryptography & secure computation

Goals
- How to prepare functions and data, such that the correct outcome can be derived from a computation which operates on protected data?

⇒ Formal definitions

Method

Proofs & implementations
- Theoretic: very strong
- Empirical: weaker

Limitations?

E.g., Fourier analysis: high crypto complexity

E.g., matrix inversion, determinants & sorting: high crypto complexity

Floating point numbers: low crypto precision

Sensor (mic)

Speech data

Feature extraction

Speech features

Classifier

Scores

Decision

Risk

Does cryptography defy decision theory? [no]
Need for taxonomies
Need for taxonomies
Need for taxonomies
Need for taxonomies

- User-centric
- Information types
- Data storage
- Speech capture
- Speech processing
Need for taxonomies

- User-centric
- Data storage
- Speech capture
- Speech processing
- Involved entities
- Information types
Need for taxonomies

- User-centric
- Data storage
- Speech capture
- Information types
- Case studies
- Speech processing
- Involved entities
Need for taxonomies

- User-centric
- Data storage
- Speech capture
- Speech processing
- Involved entities
- Safeguards
- Information types
- Case studies
Need for taxonomies
Need for taxonomies

- User-centric
- Data storage
- Speech capture
- Speech processing
- Safeguards
- Involved entities
- Case studies
- Information types

Qualitative research

Consensual (in-good-faith use of ISCA)

Unwittingly (potential ISCA abuse)

Quantitative research

(unwittingly (potential ISCA abuse))
Pre-advertisement — References

— ISCA Special Interest Group —

Security & Privacy in Speech Communication
Recent activities

- Interspeech 2019 special sessions
  - ASVspoof 2019: Future horizons in spoofed/fake audio detection
  - Privacy in Speech and Audio Interfaces

- ASRU 2019, ASVspoof follow-up

- Privacy: Speech meets legal experts

- CoSDEO: Privacy and Security in Digital Assistants
Security: free from threat or danger
Privacy: free from public attention
Established @ Interspeech 2019
75 members as of March 2020
Dissemination
- E-mail list
- www.spsc-sig.org
- LinkedIn group
- Twitter
Join us!
simply drop an email: nautsch@eurecom.fr

Security & Privacy in Speech Communication

Tom Bäckström
Chair
Andreas Nautsch
Secretary
Upcoming challenges @ Interspeech 2020

● VoicePrivacy
  ○ [https://www.voiceprivacychallenge.org](https://www.voiceprivacychallenge.org)
  ○ Can we anonymize speech to hide the biometric identity, while still recognising what was said?

● The Attacker’s Perspective on Automatic Speaker Verification
  ○ [https://sites.google.com/view/attackers-perspective-on-asv](https://sites.google.com/view/attackers-perspective-on-asv)
  ○ Which loopholes can be exploited in voice biometrics, in existing countermeasures or in both?

● Call for proposals: challenges, workshops, etc. — let’s get in touch :)