Audio Security & Privacy

Andreas Nautsch
EUROCOM

7th CoSDEO workshop

2020-03-27

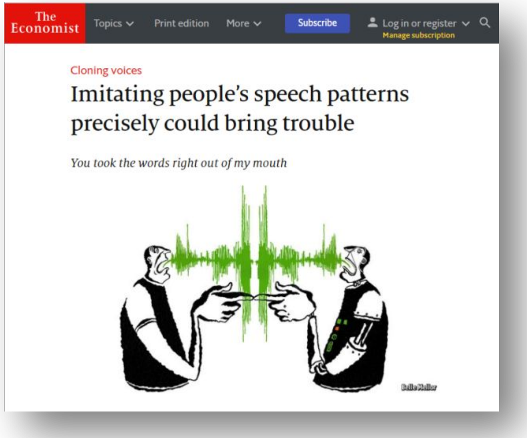
Outline

- Security: ASVspoof challenges

- Privacy: speech meets legal & crypto experts

- ISCA SIG: Security & Privacy in Speech Communication
  // ISCA: Int’l Speech Communication Association

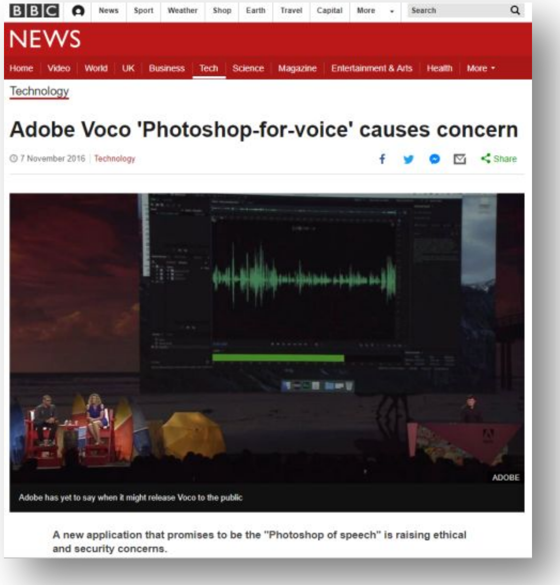
Security in voice biometrics is becoming a necessity



Cloning voices

Imitating people's speech patterns precisely could bring trouble

You took the words right out of my mouth



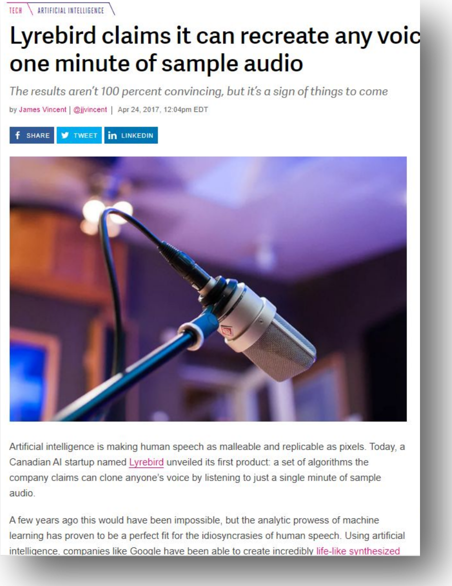
Technology

Adobe Voco 'Photoshop-for-voice' causes concern

7 November 2016 | Technology

Adobe has yet to say when it might release Voco to the public

A new application that promises to be the "Photoshop of speech" is raising ethical and security concerns.



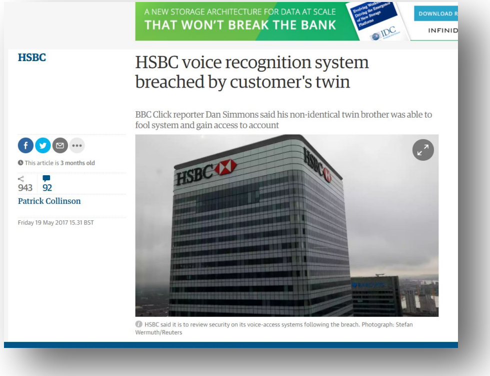
Lyrebird claims it can recreate any voic one minute of sample audio

The results aren't 100 percent convincing, but it's a sign of things to come

by James Vincent | @jjvincent | Apr 24, 2017, 12:04pm EDT

Artificial intelligence is making human speech as malleable and replicable as pixels. Today, a Canadian AI startup named Lyrebird unveiled its first product: a set of algorithms the company claims can clone anyone's voice by listening to just a single minute of sample audio.

A few years ago this would have been impossible, but the analytic prowess of machine learning has proven to be a perfect fit for the idiosyncrasies of human speech. Using artificial intelligence, companies like Google have been able to create incredibly life-like synthesized


A NEW STORAGE ARCHITECTURE FOR DATA AT SCALE
THAT WON'T BREAK THE BANK
IDC
DOWNLOAD R
INFINID
HSBC

HSBC voice recognition system breached by customer's twin

BBC Click reporter Dan Simmons said his non-identical twin brother was able to fool system and gain access to account

This article is 3 months old

943 92

Patrick Collinson

Friday 19 May 2017 15.31 BST

HSBC said it is to review security on its voice-access systems following the breach. Photograph: Stefan Wermuth/Reuters

Voice spoofing & biometric presentation attacks

greatest threats!

speech synthesis

→ text-to-speech (TTS) →

voice conversion

→ → VC →

replay

Sneakers (1992) Universal Pictures

office

mobile phone

meeting room

HQ loudspeaker

anechoic room

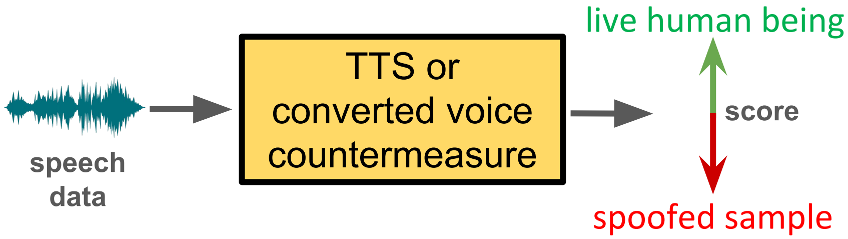
HQ loudspeaker

impersonation

Andreas here... verify my voice!

mimicry by a human being

ASVspoof 2015



speech data

TTS or converted voice countermeasure

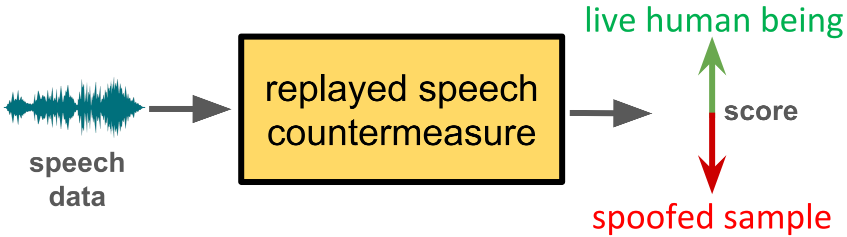
live human being

score

spoofed sample

16 organizations participated

ASVspoof 2017



speech data

replayed speech countermeasure

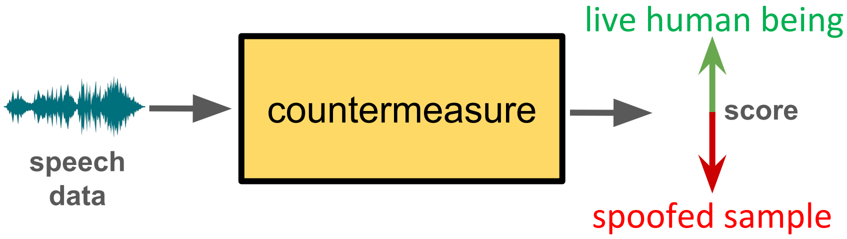
live human being

score

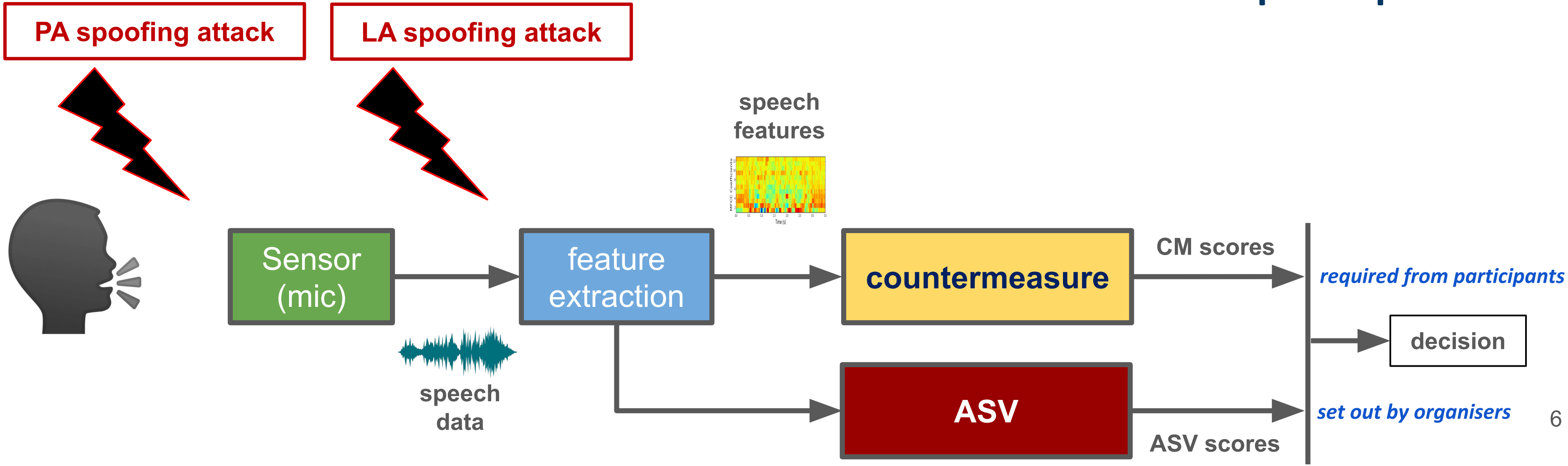
spoofed sample

49 organizations participated

ASVspoof 2019



154 teams participated



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

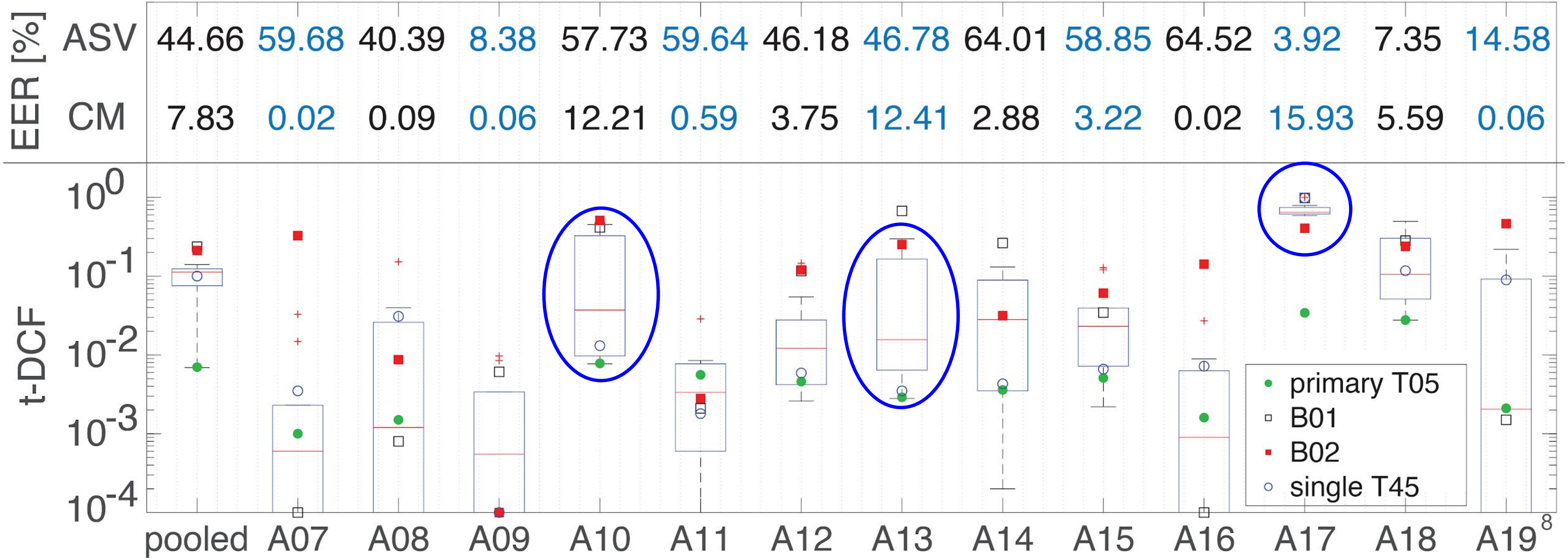
VCTK corpus

[1] C. Veaux, J. Yamagishi, K. MacDonald, "CSTR VCTK Corpus: English Multi-speaker Corpus for CSTR Voice Cloning Toolkit," University of Edinburgh. The Centre for Speech Technology Research (CSTR), 2017.

ASVspoof 2019 — Logical access attacks

13 attacks breakdown

ASV only zero-effort impostors → EER = 2.48%

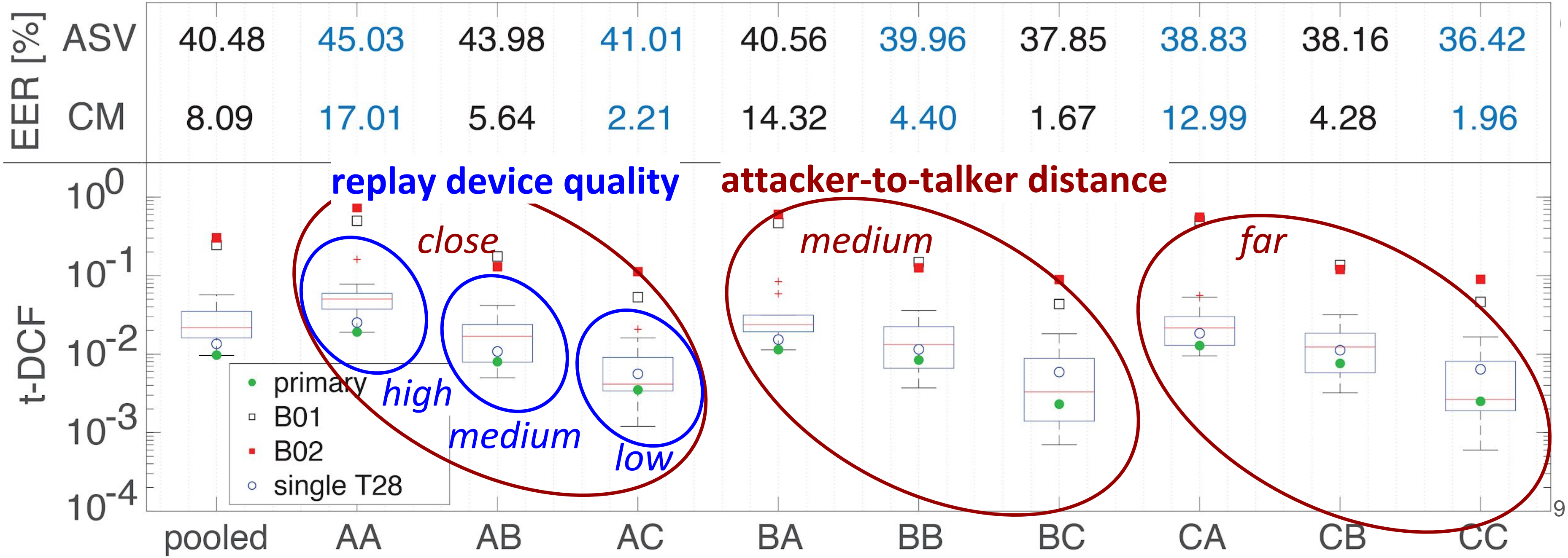


EER [%]	pooled	A07	A08	A09	A10	A11	A12	A13	A14	A15	A16	A17	A18	A19
ASV	44.66	59.68	40.39	8.38	57.73	59.64	46.18	46.78	64.01	58.85	64.52	3.92	7.35	14.58
CM	7.83	0.02	0.09	0.06	12.21	0.59	3.75	12.41	2.88	3.22	0.02	15.93	5.59	0.06

ASVspoof 2019 — Physical access attacks

9 attacks breakdown

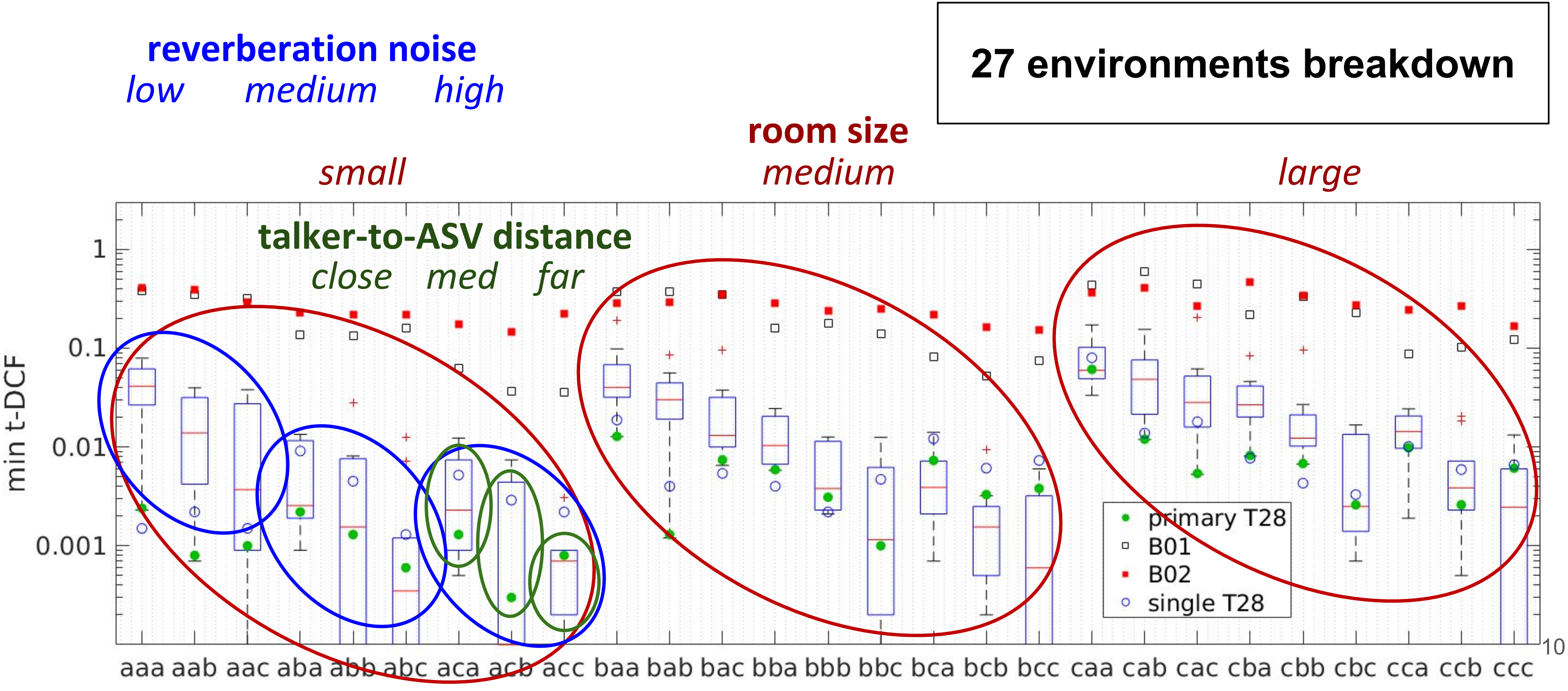
ASV only zero-effort impostors → EER = 6.47%



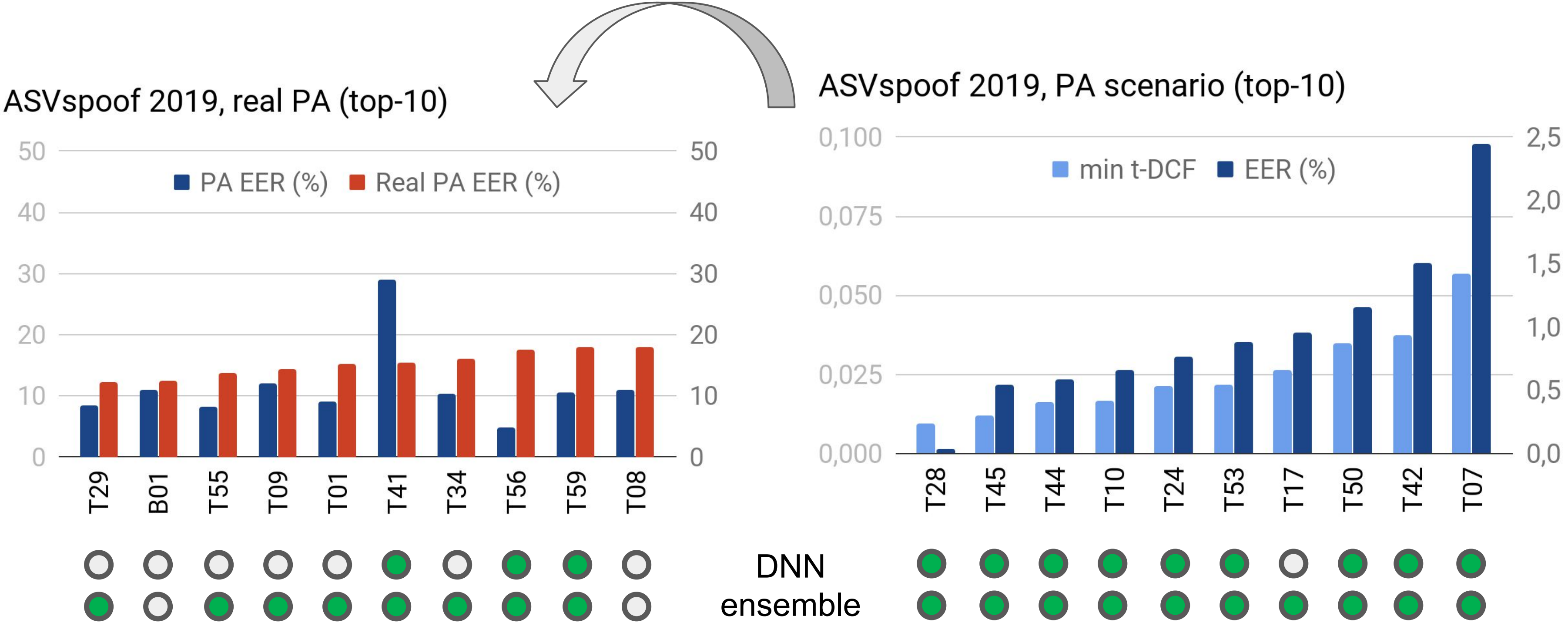
EER [%]	pooled	AA	AB	AC	BA	BB	BC	CA	CB	CC
ASV	40.48	45.03	43.98	41.01	40.56	39.96	37.85	38.83	38.16	36.42
CM	8.09	17.01	5.64	2.21	14.32	4.40	1.67	12.99	4.28	1.96

ASVspoof 2019 — Physical access attacks

27 environments breakdown



ASVspoof 2019 — “the hidden track of the album”



ASVspoof 2019, real PA (top-10)

ASVspoof 2019, PA scenario (top-10)

ASVspoof 2019 — Organisers

Junichi Yamagishi

NII, Japan
Univ. of Edinburgh, UK

Massimiliano Todisco

EURECOM, France

Md Sahidullah

Inria, France

Héctor Delgado

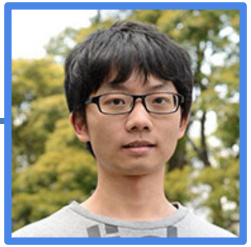
EURECOM, France
Nuance, Spain

Nicholas Evans

EURECOM, France

Ville Vestman

UEF, Finland



Xin Wang

NII, Japan

Tomi H. Kinnunen

UEF, Finland

Kong Aik Lee

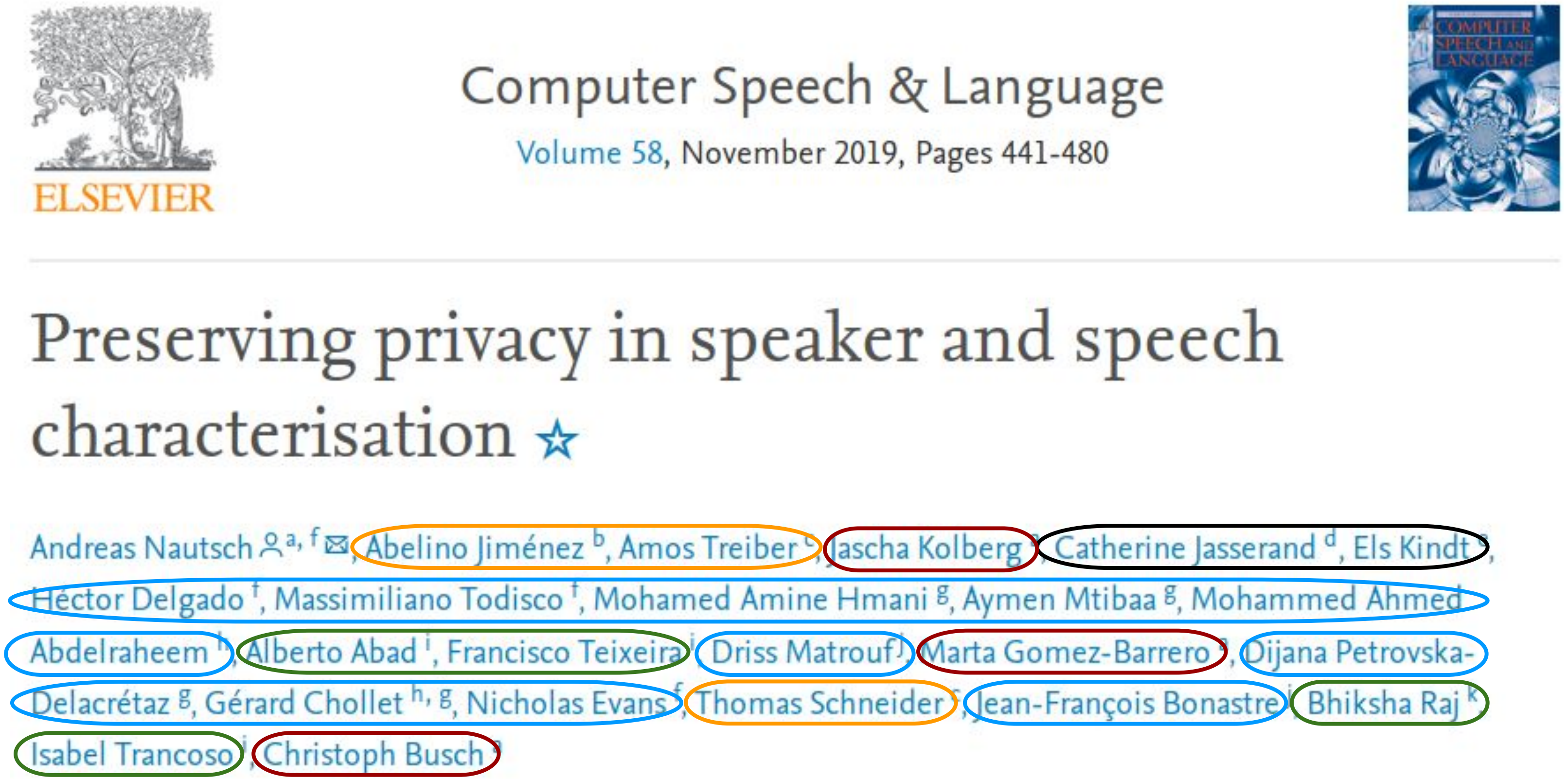
NEC, Japan

Andreas Nautsch

EURECOM, France

— Privacy —

speech meets legal & crypto experts



Computer Speech & Language

Volume 58, November 2019, Pages 441-480

Preserving privacy in speaker and speech characterisation ☆

Andreas Nautsch [a, f], Abelino Jiménez [b], Amos Treiber [c], Jascha Kolberg [e], Catherine Jasserand [d], Els Kindt [e], Héctor Delgado [f], Massimiliano Todisco [f], Mohamed Amine Hmani [g], Aymen Mtibaa [g], Mohammed Ahmed Abdelraheem [h], Alberto Abad [i], Francisco Teixeira [i], Driss Matrouf [j], Marta Gomez-Barrero [e], Dijana Petrovska-Delacrétaz [g], Gérard Chollet [h, g], Nicholas Evans [f], Thomas Schneider [c], Jean-François Bonastre [j], Bhiksha Raj [k], Isabel Trancoso [i], Christoph Busch [e]

Speaker recognition

Study of the Law

Biometrics

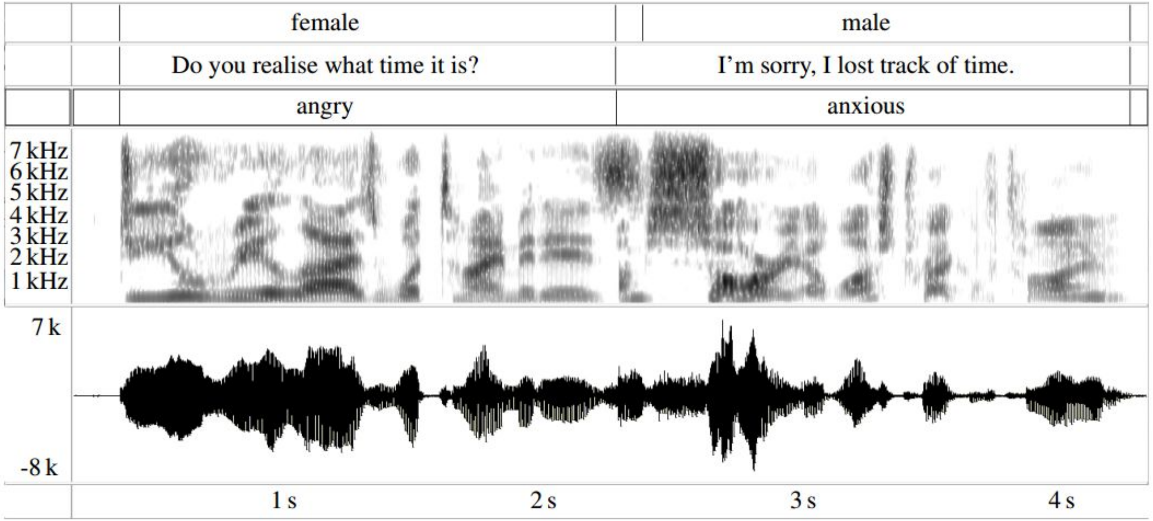
Speech communication

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

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

- 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*

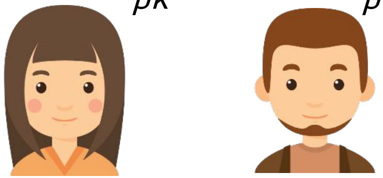
  - *Intel SGX: hardware-assisted security*

- How to check, we did well?
- Spoiler: always have a crypto expert around — plenty of space for mistakes

Privacy & speech data; cybersecurity II/III

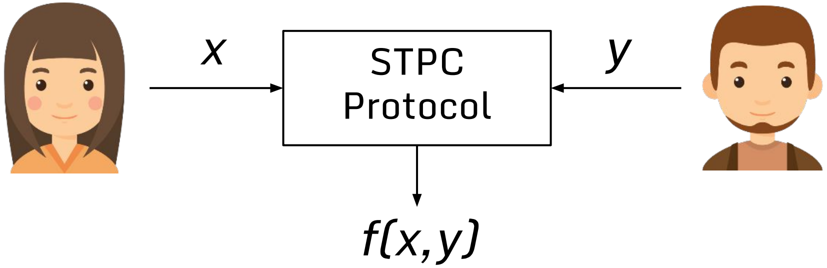
- Odyssey 2018: HE for speaker recognition

$$Enc_{pk}(x) \boxplus Enc_{pk}(y) = Enc_{pk}(x+y)$$



"compute all-at-once" approach
Slow computation
Low communication

- Interspeech 2019: HE & STPC
- Speech Communication 2020: STPC



x

STPC Protocol

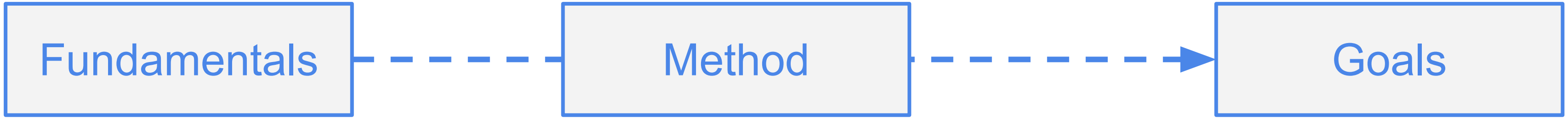
y

$f(x,y)$

"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



Assumptions

Cryptography & secure computation

Proofs & implementations

Theoretic: very strong
Empirical: weaker

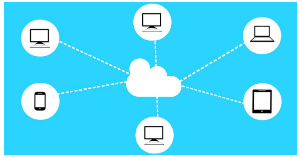
How to prepare functions and data, such that the correct outcome can be derived from a computation which operates on protected data?

⇒ Formal definitions

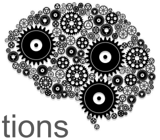
Computational indistinguishability

Cryptographic hardness

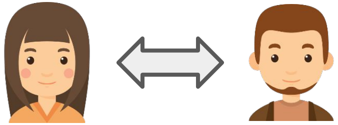
Zero knowledge
“zero evidence”



Infrastructure

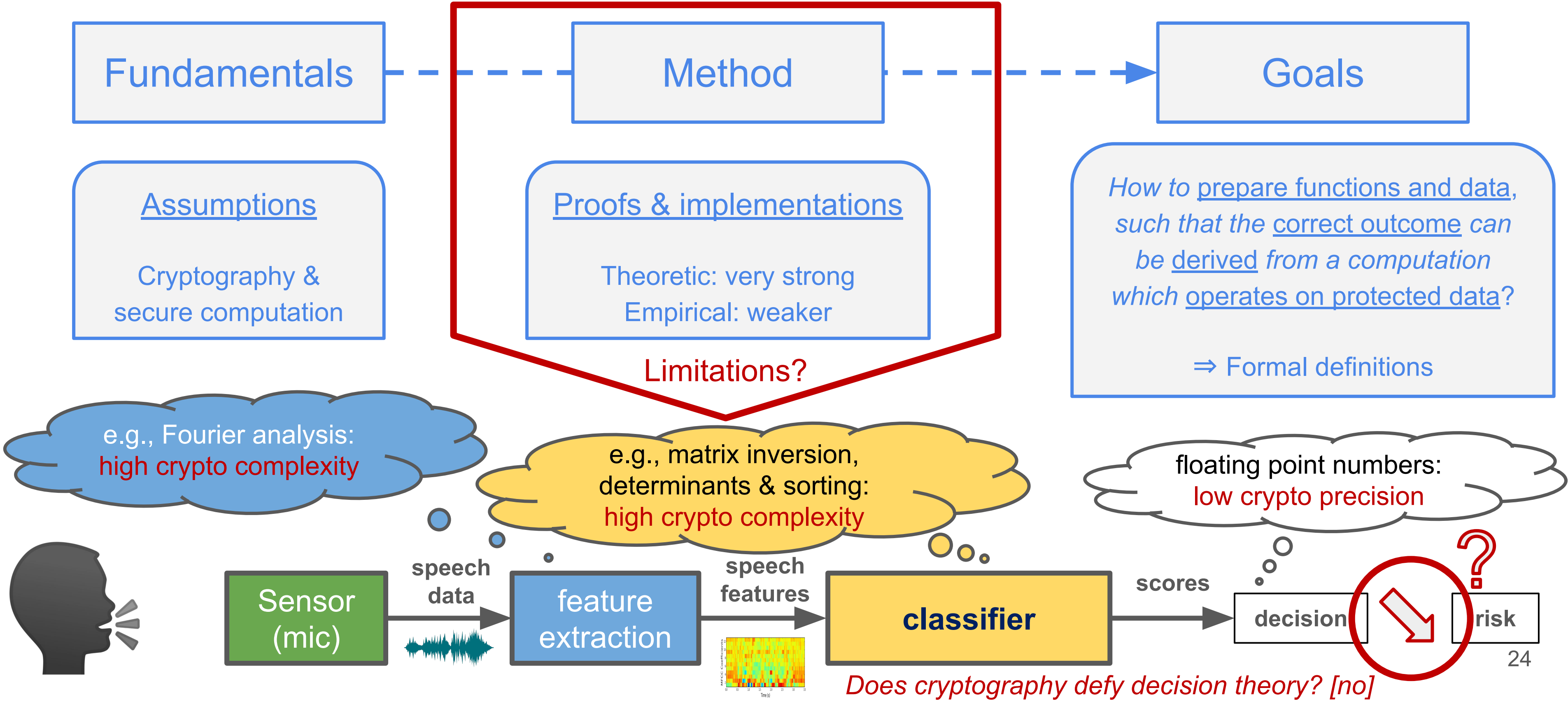


Computations



Communication

Privacy & speech data; cybersecurity — easy, right?



Fundamentals

Method

Goals

Assumptions

Cryptography & secure computation

Proofs & implementations

Theoretic: very strong
Empirical: weaker

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

⇒ Formal definitions

Limitations?

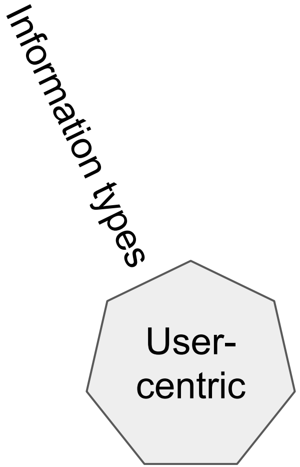
e.g., Fourier analysis:
high crypto complexity

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

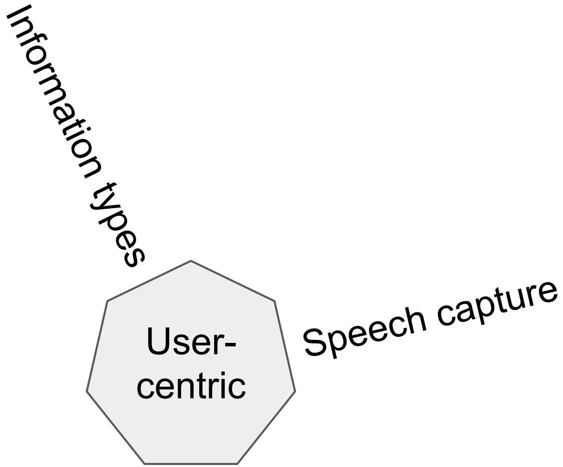
floating point numbers:
low crypto precision

Does cryptography defy decision theory? [no]

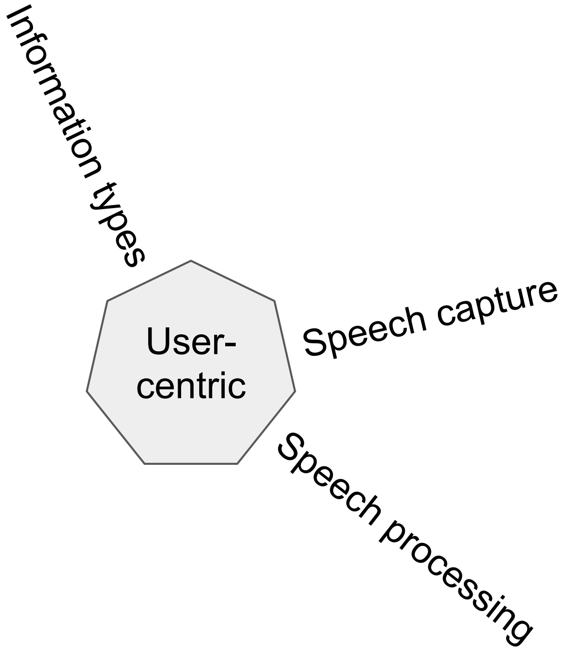
Need for taxonomies



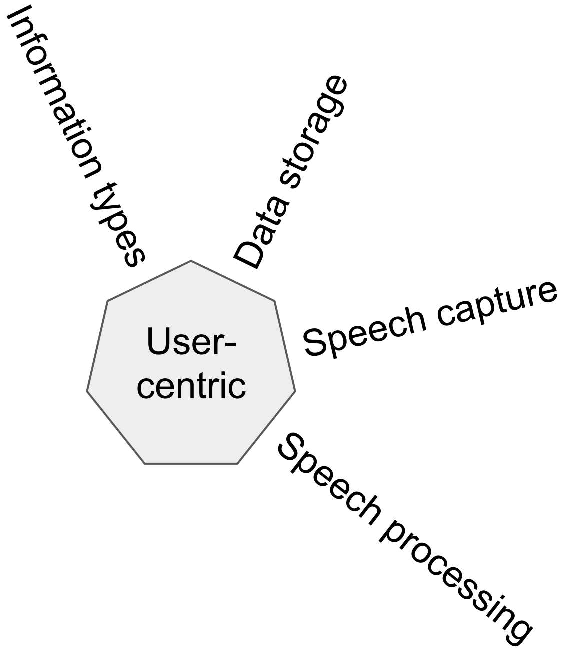
Need for taxonomies



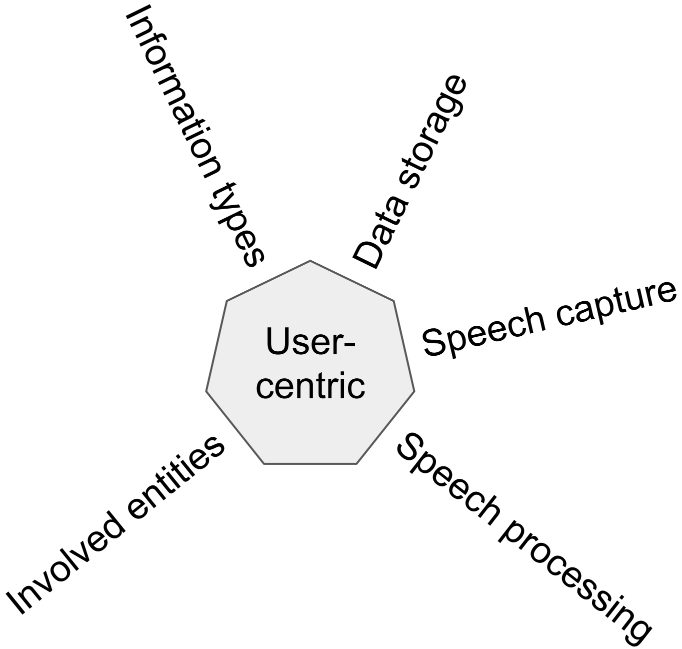
Need for taxonomies



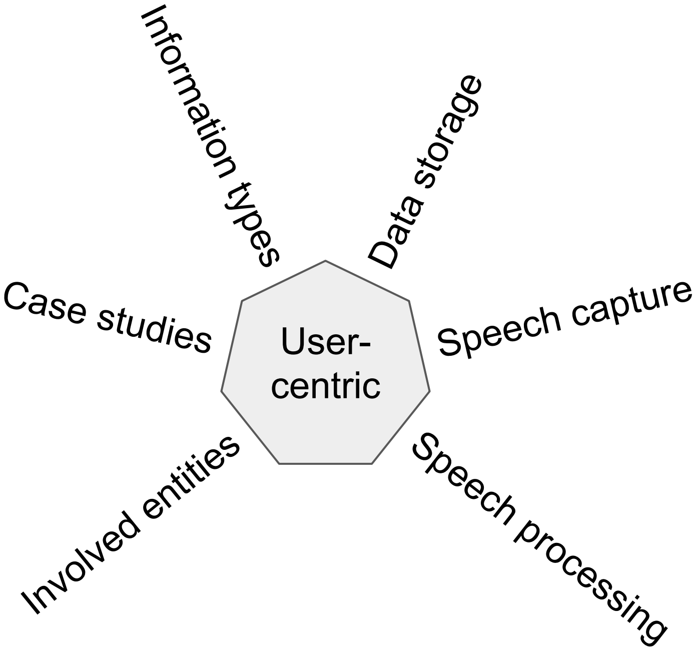
Need for taxonomies



Need for taxonomies



Need for taxonomies



User-centric

Information types

Data storage

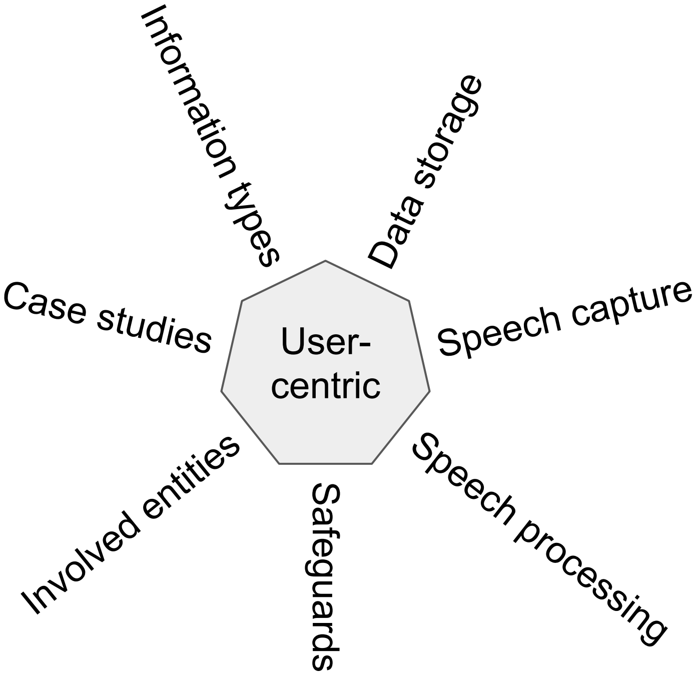
Case studies

Speech capture

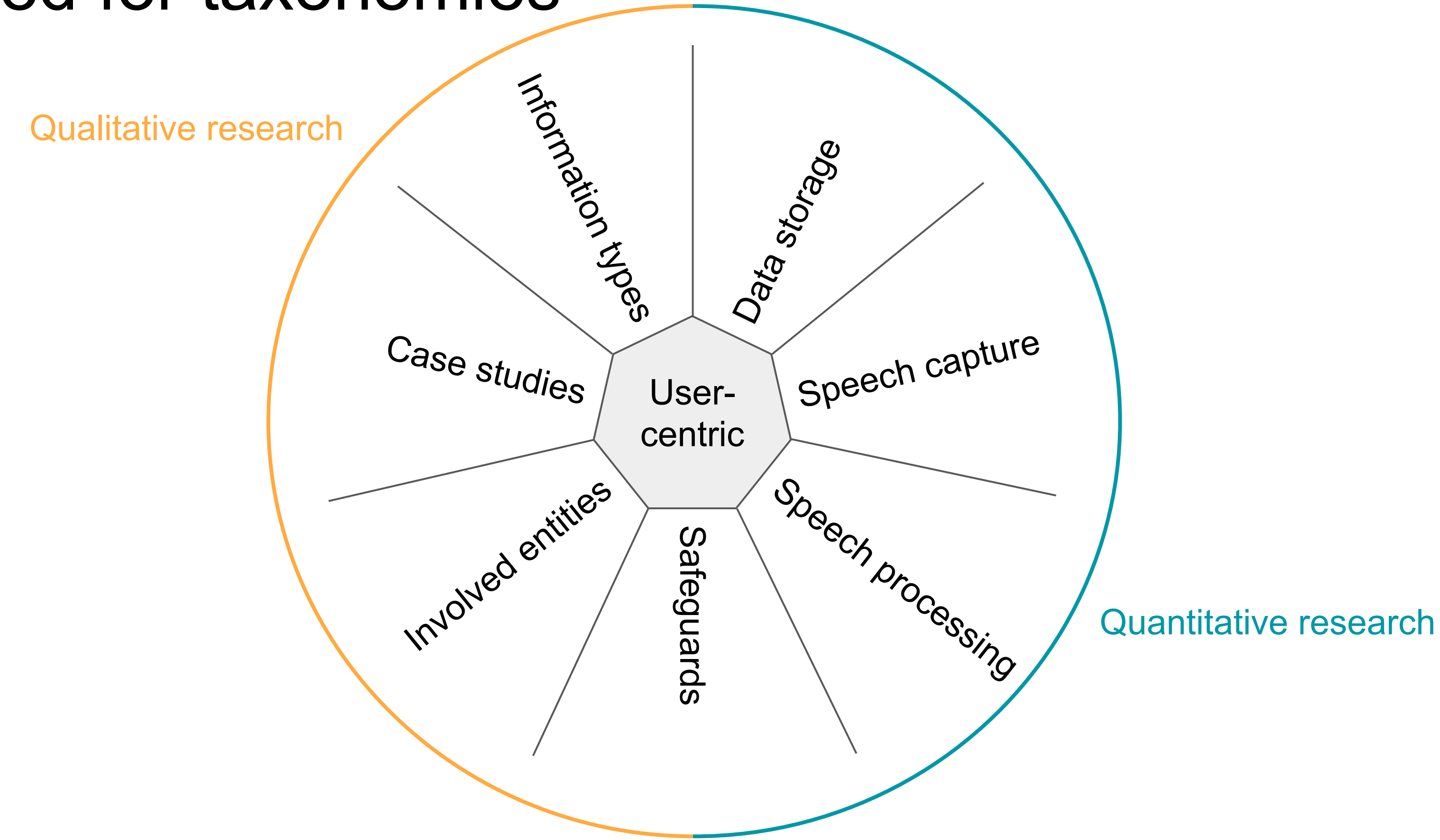
Involved entities

Speech processing

Need for taxonomies



Need for taxonomies



Qualitative research

Information types

Data storage

Case studies

User-centric

Speech capture

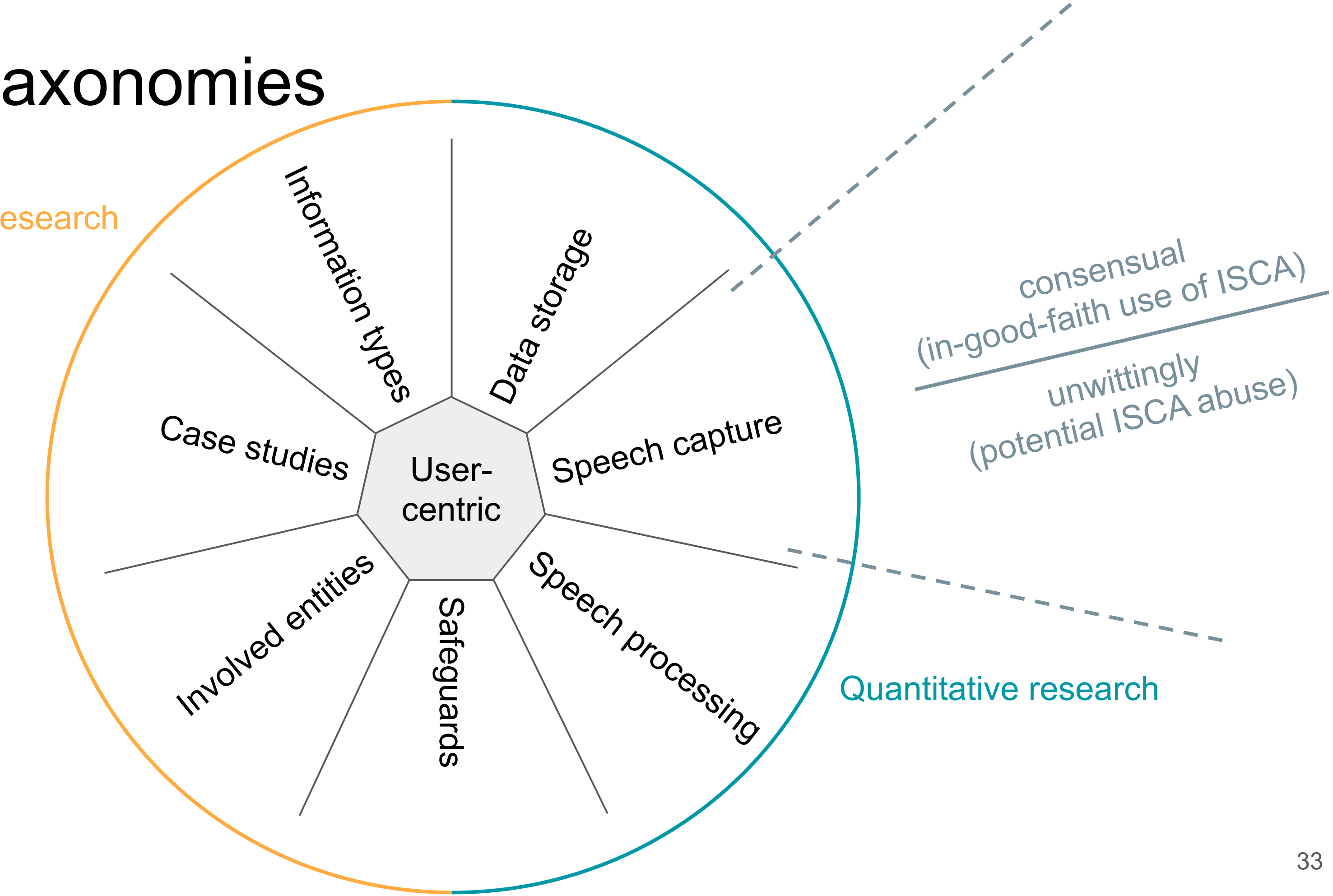
Involved entities

Safeguards

Speech processing

Quantitative research

Need for taxonomies



Pre-advertisement — References

[1] **Todisco et al.** "ASVspoof 2019: Future horizons in spoofed and fake audio detection," Proc. Interspeech, 2019

[2] **Wang et al.** "The ASVspoof 2019 database," Computer Speech & Language, to appear

[3] **Kinnunen et al.** "Tandem Assessment of Spoofing Countermeasures and Automatic Speaker Verification: Fundamentals," IEEE/ACM-T-ASLP, submitted

[4] **Nautsch et al.** "Preserving privacy in speaker and speech characterisation," Computer Speech & Language, Vol.58, November 2019

[5] **Nautsch et al.** "The GDPR & speech data: Reflections of legal and technology communities, first steps towards a common understanding," Proc. Interspeech, 2019

[6] **Nautsch et al.** "Homomorphic encryption for speaker recognition: Protection of biometric templates and vendor model parameters," Proc. Odyssey, 2018

[7] **Nautsch et al.** "Privacy-preserving speaker recognition with cohort score normalisation," Proc. Interspeech, 2019

[8] **Treiber et al.** "Privacy-preserving PLDA speaker verification using outsourced secure computation," Speech Communication, Vol.114, 2019

[9] **Bayerl et al.** "Privacy-preserving speech processing via STPC and TEEs," Proc. Privacy Preserving Machine Learning, CCS Workshop, 2019

[10] **Schneider and Treiber** "A Comment on Privacy-Preserving Scalar Product Protocols as proposed in 'SPOC'," IEEE Transactions on Parallel and Distributed Systems, Vol. 31(3), 2020

[11] **Kröger et al.** "Privacy Implications of Voice and Speech Analysis – Information Disclosure by Inference," Proc. IFIP Summer School, Springer Privacy and Identity Management, Data for Better Living: AI and Privacy, 2020

— ISCA Special Interest Group —

Security & Privacy
in Speech Communication

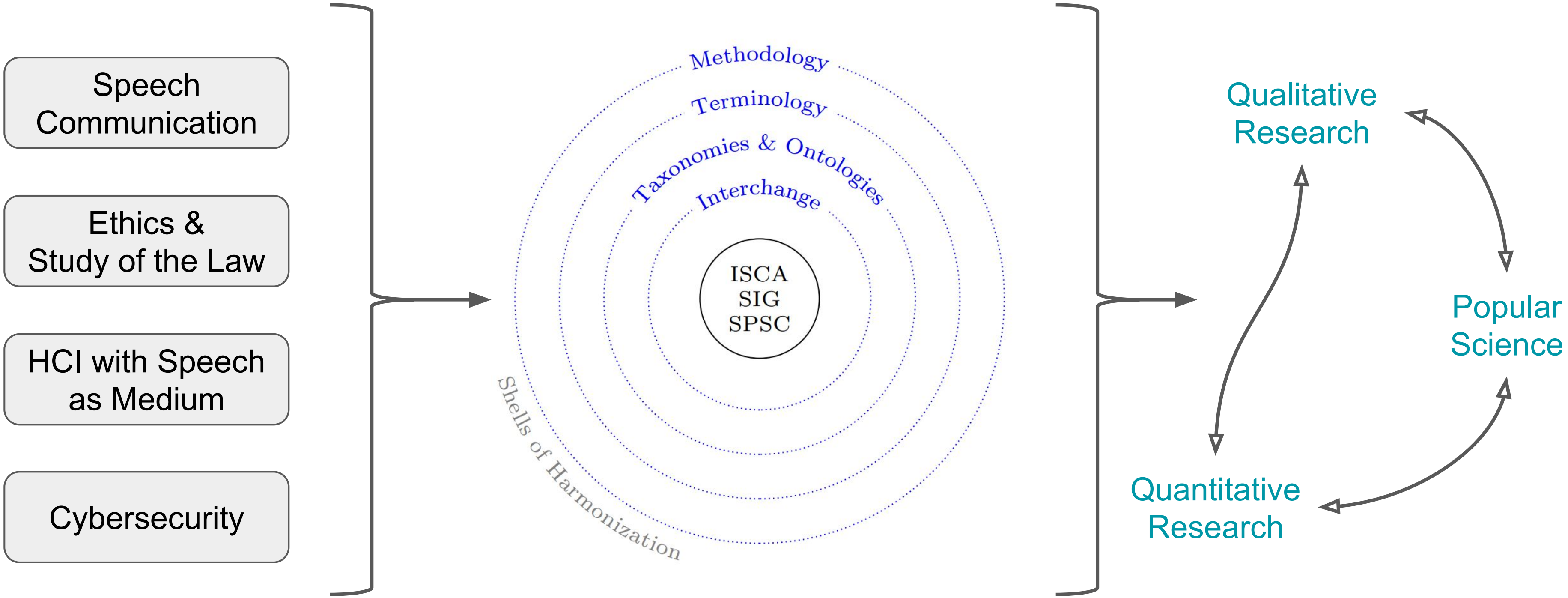
Recent activities

- Interspeech 2019 special sessions *September 2019*
    - ASVspoof 2019: Future horizons in spoofed/fake audio detection
    - Privacy in Speech and Audio Interfaces
- ASRU 2019, ASVspoof follow-up *December 2019*
- Privacy: Speech meets legal experts *January 2020*
- CoSDEO: Privacy and Security in Digital Assistants *March 2020*

Research Areas

Understanding Security & Privacy

Dissemination



Speech Communication

Ethics & Study of the Law

HCI with Speech as Medium

Cybersecurity

Qualitative Research

Popular Science

Quantitative Research

Security: free from threat or danger

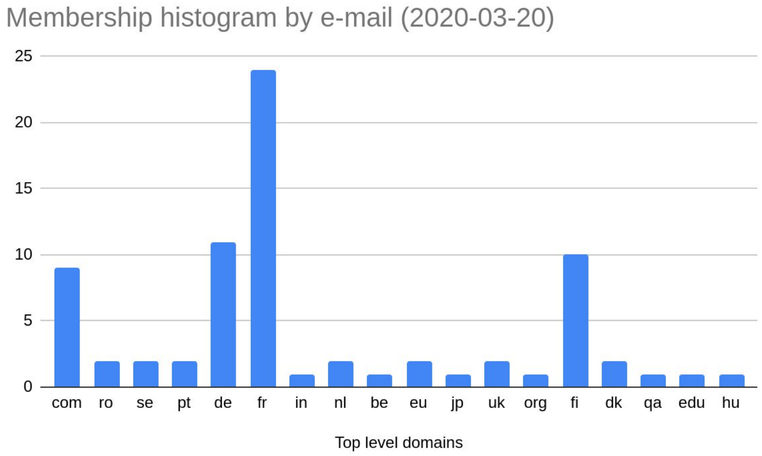
Privacy: free from public attention

Security & Privacy in Speech Communication

- 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



Tom Bäckström

Chair

Andreas Nautsch

Secretary

Upcoming challenges @ Interspeech 2020

- VoicePrivacy
    - 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
    - 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 :)