Aalto University

CS-E4070 – Special Course in Machine Learning and Data Science:

Privacy in speech and audio interfaces

Introduction

Tom Bäckström

Department of Signal Processing and Acoustics
Aalto University, School of Electrical Engineering
tom.backstrom@aalto.fi

22.01.2019 – 02.04.2019

Welcome

Program today

- Hello! $\sim$ 10 min
- Introduction to “privacy in speech and audio interfaces” $\sim$ 60 min
- Practical organization $\sim$ 20 min

Introduction to

Privacy in speech and audio interfaces

Dominant trend in Speech
Speech Interfaces and IoT

Personal digital assistants
and speech interfaces everywhere!

- Homepods and home-automation
- Smartphones, -TVs, -watches
- Home appliances
- ..

Smart speaker (homepod) sales ∼11.7M in 2018/Q2
(Source: Strategy Analytics Smart Speaker service).

Benefits
of Speech Interfaces with IoT

- Speech is our most natural form of communication.
  - Speech is natural also in human-to-computer communication.
- Access any nearby device.
  - No need to find device. (*Where's my phone?*)
- Shift focus *from device to task*.
  - Suppose I want to watch a movie.
  - Searching my phone or remote is a distraction.
- Single hands-free interface to everything.

Challenges with current paradigm
Scenario I: Cafeteria

- Alice and Bob meet at a cafeteria.
  - Alice uses Apple.
  - Bob uses Google.
  - Cafeteria offers Amazon/Alexa.
- Which interface should they use?

  $\Rightarrow$ Awkward user experience.

- Products are *not interoperable*.

  $\Rightarrow$ Inefficient use of resources.
  - Cannot use all hardware.
  - Cannot use all services.

$\Rightarrow$ Not a good UI.

Photo by Valeria Boltneva from Pexels

Outlook in current paradigm
Scenario I: Cafeteria

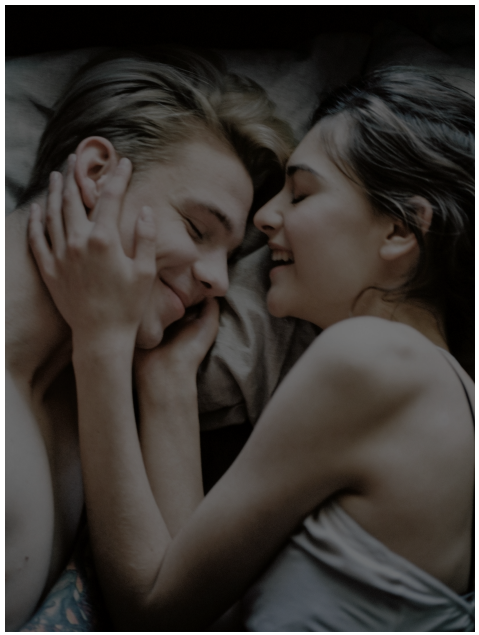
- Available products are great in isolation.
- However, multi-device scenarios are currently unsatisfactory.

$\Rightarrow$ We need collaboration and interoperability.

$=$ An open API between devices to manage speech commands.

Challenges with current paradigm
Scenario II: Bedroom

- "*Computer, lights off.*"
- *No* outsider may *ever* access my bedroom.



Creative Commons, Photo from Pexels

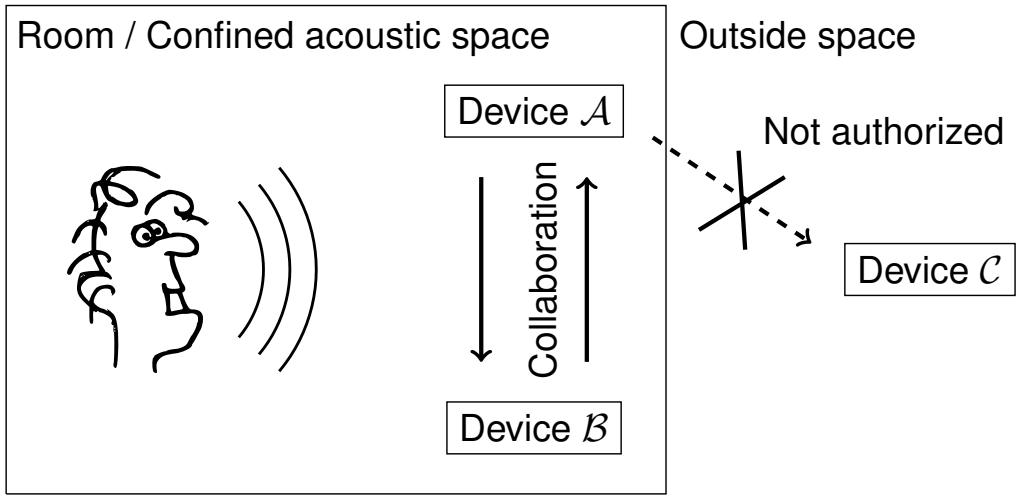
- UI-design problem
  - Controlling, reporting and automatically adapting to privacy.
- Systems-design problem
  - Optimize privacy (not "preserve")
- Enineering problem
  - Determine desired level of privacy.
  - Enforce privacy.

Challenges with current paradigm

Contradicting objectives?

- More collaboration for better UI.
- More privacy for better UI.

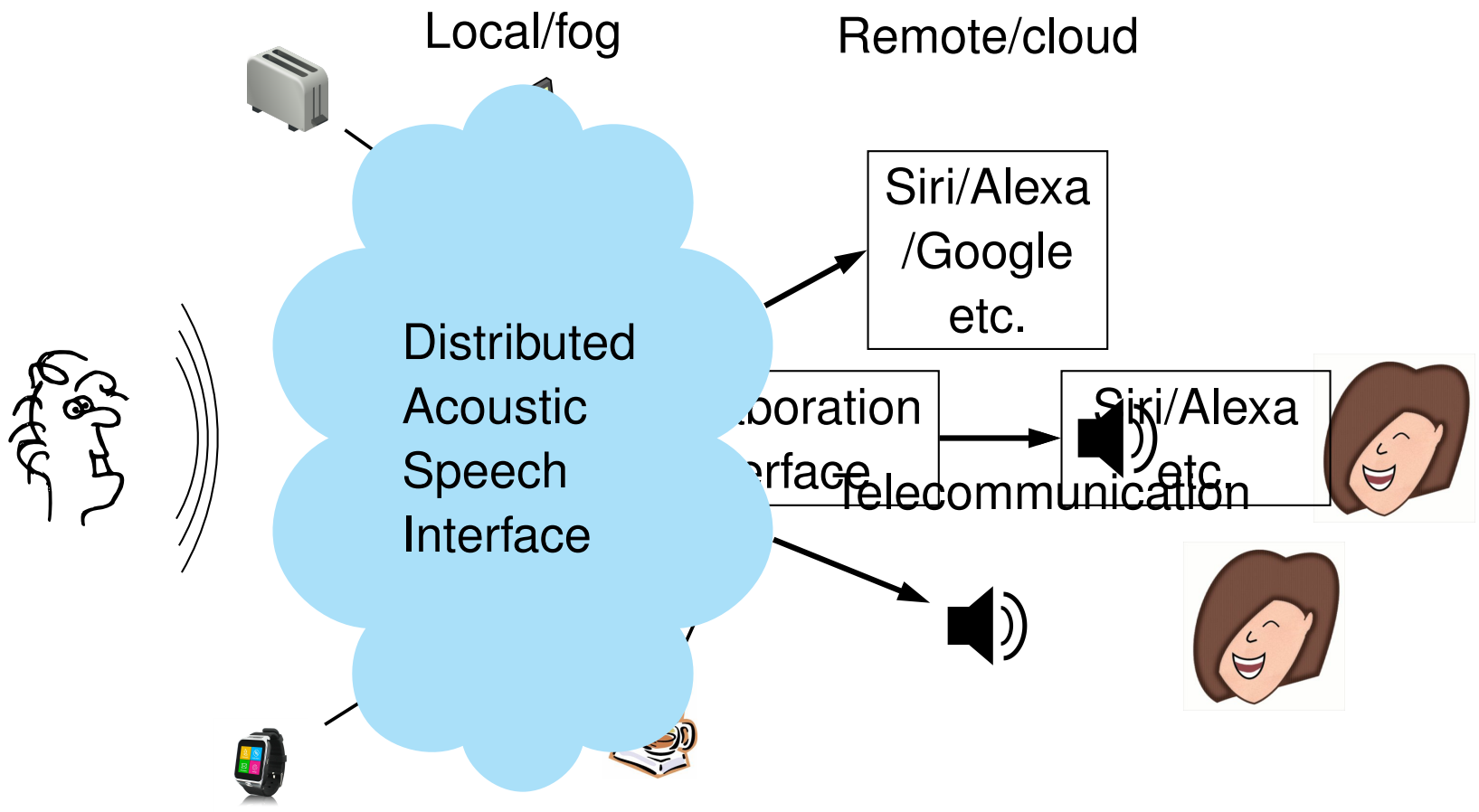
Solution: Local collaboration



- Local devices can hear the sound anyway.
  - Local collaboration does not break privacy.
    = Wireless acoustic sensor network (WASN).
- Privacy is an issue only with *remote* participants.
  - Remote in space or time.

- Simple tasks can be handled locally.
  - *"Computer, lights off."*
- Only complex tasks and specific requests require access to remote participants/devices.

Solution: Acoustic collaboration interface



Solution: Acoustic collaboration interface

Requirements:

- Communication of acoustic data = distributed speech coding
- Authorization based on neighborhood = acoustic fingerprinting
- Transparent and interoperable = need standard?
  - Allows third-party auditing of privacy.

- Lots of devices = must be low-complexity.
- Multi-microphone source separation/enhancement at decoder.

Conclusion

$\Rightarrow$ We need an open API between speech interfaces.

- For device-to-device interfaces in person-to-person and person-to-service communication.
- High-quality PDAs as cloud service
  - *“Siri, how’s the weather in Japan?”*
  - *“Alexa, play me the movie Batman returns?”*
- Allows small companies to produce new innovative device.
  - E.g. Sonos can focus on their core-business, high-quality audio, but give access to Alexa/Google/Siri.

Research projects

Sound Privacy
Research Project

- What is privacy? How much privacy does the user want?
- Hypothesis:
  - Perception of privacy depends on the acoustic environment.
  - We can analyze the acoustic envirnoment to determine desired level of privacy.
- Approach:
  - Record dialogues in different environments (acoustic recording).
  - On-site questionnaire for participants: "*Could you share a secret in this environment?*".
  - Suplement recordings with artificial speech+noise+RIR signals.
  - On-line questionnaire where subjects hear recorded dialogues and answer the same questions.

Acoustic Fingerprinting
Research Project

- Are these devices in the same room?
  Is collaboration allowed?
- Hypothesis:
  - Devices in the same room have similar microphone signals.
  - Devices which hear the same signal are allowed to collaborate.
- Approach:
  - Generate fingerprint from acoustic signal (with error correction).
  - Use fingerprint as cryptograhpic key for inter-device communication.

Dynamic Access Management for Distributed Speech Interfaces

Research Project

- Mobile devices come and go (ad-hoc collaboration).
  Access management needs to take that into account.
- Hypothesis:
  - Distributed processing requires distributed access management.
    For example, resource optimization requires knowledge of nodes.
- Approach:
  - Distributed database of access rights.
  - Dynamically granting and revoking access.
  - Must be immutable and distributed.
  - Pick&choose suitable components from distributed ledger (i.e. block-chain) technologies.

Unified distributed acoustic interface for person-to-person and person-to-service

Research Project

- In communication with a remote person and with a cloud server we are transmitting the same speech signal.
- Hypothesis:
  - It would be efficient to use the same codec for both purposes.
  - Collaboration between devices (distributed coding) improves efficiency.
- Approach:
  - Design coding with conventional codecs as baseline $\Rightarrow$ preserve single channel quality.
  - Incorporate elements of distributed coding methods.
  - Low-complexity encoders, intelligent decoder.

Details

Resource Optimization for Speech Interfaces
Research Project

- IoT devices often have limited resources.
- Balance between local CPU-use and transmission bandwidth for optimal energy efficiency?
- Hypothesis:
  - Overall resources can be optimized by adjusting trade-off between bandwidth and CPU-load.
- Approach:
  - Integrate load optimization into distributed transmission/coding system.
  - Each devices chooses its own operations based on a local energy optimization function.

Questions?

Opportunities

More for everyone!

- User gets better UI – private, intuitive and high quality audio
  - Unified voice services; UI and telecom
- Personal digital assistants as a cloud service from any device:
  - *“Siri, how’s the weather in Japan?”*
  - *“Alexa, play me the movie Batman returns?”*
- Ecosystem for small and large companies.

- Speech coding is again important!
  - Need efficient communication between IoT devices.

... but machine learning ...

- Machine learning relies on “big data”.
  - “We need data to train our models” = Ends justify the means?
- Counter-example:
  - Fact: Releasing all medical data would enable great innovations and save lives.
  - But we still prefer to keep medical data private.
- This is not someone elses’ problem.
  - You are responsible.

Afterthoughts

- Privacy is a dimension of the quality of the user-interface.
  - An intuitive UI is a good UI.
  - Privacy is not binary (yes/no), but a grayscale; *Would I tell my mom/neighbour/colleague?*
- Speech is about dialogue.
  - Ownership of dialogue is shared among those present.
  - The EULA of one user cannot cover other users.
- A large cloud is a large target for criminals and foreign spies.
  - Distributed and heterogeneous is more resilient
    = consequences of a breach are limited.
- As a measure of privacy, replace “cloud” by a person:
  - Greg follows your every move and stores all your queries.

  $\Rightarrow$ Greg is a $<$XXXXXXXXXX$>$?

Practical Organization

Practical organization

- Main teacher:
  Tom Bäckström mailto:tom.backstrom@aalto.fi
- Teaching assistants:
  Sneha Das mailto:sneha.das@aalto.fi
  Pablo Pérez Zarazaga mailto:pablo.perezzarazaga@aalto.fi

Schedule

22.1.2019	14-16	Practical organization. Introduction to privacy in speech and audio interfaces.
29.1.2019	10-12 and 14-16	Visiting lectures by prof Susanna Lindroos-Hovinheimo (U Helsinki), prof Stephan Sigg (Aalto) and Docent Michael Laakasuo (U Helsinki).
5.2.2019	14-16	Selecting topics for seminar projects, reports and presentations. Set up work schedules.
10.2.2019	noon	DL for experimental project proposals (5 ECTS version; send to teacher by email)
19.3.2019	14-16	Seminar presentations
26.3.2019	14–16	Seminar presentations + final discussion. DL for the first versions of the reports (for feedback).
24.4.2019	noon	Final DL for reports.

Work-load and requirements

For 5 ECTS

Participation	$\sim 6 \times 2$ h	= 12 h
Presentation	$\sim 0.5 \ldots 1$ weeks	= 20 ... 40 h
Report	$\sim 0.5 \ldots 1$ weeks	= 20 ... 40 h
Project	$\sim \quad 1 \ldots 2$ weeks	= 40 ... 80 h
Total		= 133 h

For 3 ECTS

Participation	$\sim 6 \times 2$ h	= 12 h
Presentation	$\sim 0.5 \ldots 1$ weeks	= 20 ... 40 h
Report	$\sim 0.5 \ldots 1$ weeks	= 20 ... 40 h
Total		= 80.1 h

Presentation

- Teaching objective: Practice 1. presentation, 2. summarization of material, and 3. in-depth learning of narrow topic.
- Length: XX min + 5 min discussion
- Objective of presentation: Give an introduction and overview in the topic chosen = not just copying content of the related article.
- Target audience: Generally knowledgeable engineers and engineering students
- Use the Aalto presentation template
  - Latex `https://version.aalto.fi/gitlab/latex/aaltobeamer`
  - PowerPoint `https://materialbank.aalto.fi`

Report option 1

- Task: Technical report in the scientific format
- Teaching objective: Practice 1. technical writing, 2. summarization of material, and 3. in-depth learning of narrow topic.
- Length: Max 4 pages text + 1 page references
- Objective of report: Give an introduction and overview in the topic chosen = not just copying content of the related article.
- Target audience: Generally knowledgeable engineers and engineering students
- Template: `https://www.ieee.org/conferences/publishing/templates.html`

Report option 2 (NEW)

- Task: Blog-post or magazine article for the general public
- Teaching objective: Practice 1. popular-science writing, 2. summarization of material, and 3. in-depth learning of narrow topic.
- Length: 5 ... 10 min read
- Objective of report: Give an introduction and overview in the topic chosen to the general public
- Target audience: Generally knowledgeable *non-engineers*
- Model to go by
  `https://uxdesign.cc/whisper-your-secrets-59753e04634b?source=friends_link&sk=e7320b83837ddd1bf62e602d914573ab`

Project

Can be one of the following

- Implementation and experiments within the topic area
- Literature review
- Popular-science text
  (report must then be in the scientific format)

Students choose the approach (teacher approves).

Choice of topics

- Either
  - choose topic/article from the homepage (mycourses) or
  - suggest your own topic which combines privacy with speech/audio.
- On the 5.2. lecture, topics are chosen/assigned to students.
- To avoid overlapping preferences, if you do not have your own topic, please select at least 2 topics from the web-page which you would be comfortable with.
- If there are a lot of students, then we will add more papers to the list before 29.2.
- Each student will do a work-plan + schedule on 29.2 to match his topic (instructions provided later).

That’s all for today!