REATA: A Privacy-Preserving Vocal Personal Digital Assistant

Gérard Chollet (CNRS-SAMOVAR)
Fathy Yassa (CEO SpeechMorphing)
Nigel Cannings (CTO Intelligent Voice)

gerard.chollet@telecom-sudparis.eu

Intelligent Voice®

Une école Institut Mines-Télécom

TELECOM SudParis

cnrs Depuis 80 ans, nos connaissanc bâtissent de nouveaux mondes

e-VITA EU-JAPAN VIRTUAL COACH FOR SMART AGEING

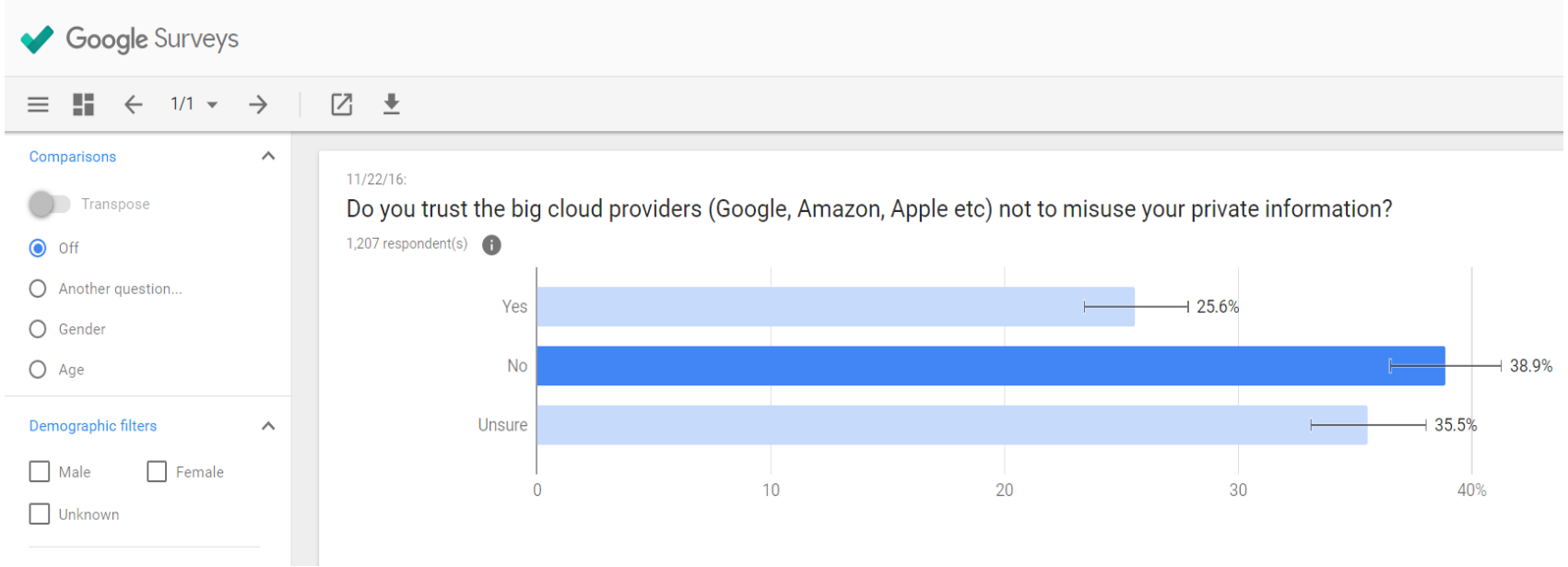
speechmorphing

Some Issues with Cloud Computing

- Siri, Alexa, Cortana, Google Now (or a hacker who breaks into them) can
  - Use (edit) your voice recordings to impersonate you (voice cloning)
  - Learn about you
    - Your identity, gender, nationality (accent), emotional state, personal data,...
  - Track you from uploads / communications of voice recordings
- *Not a futuristic scenario*
  - *Everytime you use your voice, you leave a print behind!!*

Opinions

How happy are we about living our lives in the thrall of the big harvesters of data, the Googles, the Amazons, the Apples?



Source:
We live in the Big Cloud: And we hate it… Is it time for Hipster IT?
Nigel Cannings, CTO Intelligent Voice
https://hackernoon.com/we-live-in-the-big-cloud-and-we-hate-it-is-it-time-for-hipster-it-1f130a44d2b8#.tqb3xnsdl
https://surveys.google.com/reporting/survey?survey=fwn6wwimqoqkezlhur3zrdh2oe

The Legislation

General Data Protection Regulation

GDPR.EU

Personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes

The controller shall be able to demonstrate that the data subject has consented to processing of his or her personal data

CPRA
California Privacy Rights Act of 2020

On-device Computing

FANS: Fusing ASR and NLU for on-device SLU

Martin Radfar, Athanasios Mouchtaris, Siegfried Kunzmann, and Ariya Rastrow

Alexa Machine Learning, Amazon

Extreme Model Compression for On-device Natural Language Understanding

Kanthashree Mysore Sathyendra Amazon Alexa ksathyen@amazon.com

Samridhi Choudhary Amazon Alexa samridhc@amazon.com

Leah Nicolich-Henkin Amazon Alexa nicolich@amazon.com

ATTENTION BASED ON-DEVICE STREAMING SPEECH RECOGNITION WITH LARGE SPEECH CORPUS

Kwangyoun Kim, Kyungmin Lee*, Dhananjaya Gowda, Junmo Park, Sungsoo Kim, Sichen Jin, Young-Yoon Lee, Jinsu Yeo, Daehyun Kim, Seokyeong Jung, Jungin Lee, Myoungji Han, Chanwoo Kim*

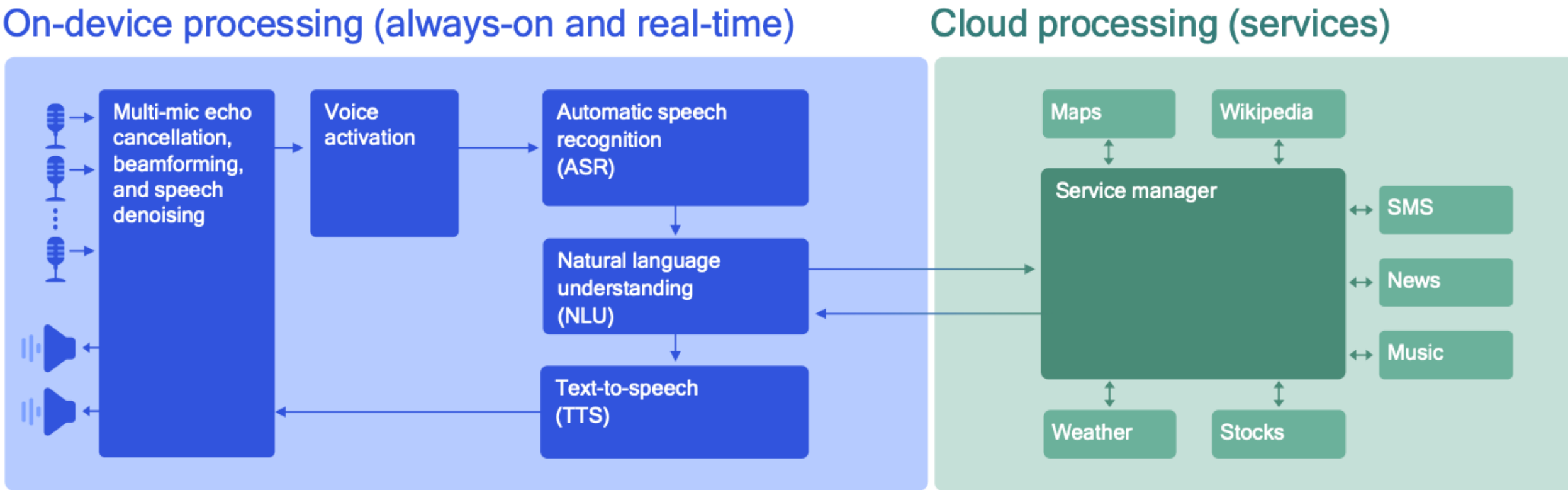
Samsung Electronics Co., Ltd., Korea

On-device vs Cloud vs Hybrid

- Our SmartPhones have enough computing power to achieve acceptable speech recognition and understanding performance,
- Although initial training could be done in the cloud, the models can be adapted on-device to the owner’s voice, attitudes and habits,
- Previous conversations can be stored on the SmartPhone as knowledge graphs to help understanding and timely responses,
- Speaker verification could be used so that REATA is well aware of who she is talking to,
- Anonymisation of the speech of the owner is proposed to make sure that the identity of the speaker is not revealed in case a cloud based system is used,
- REATA can interogate web services upon requests from the owner,
- Speech synthesis is performed on-device using the preferred voice of the owner,

Moving voice UI functionality to the end device

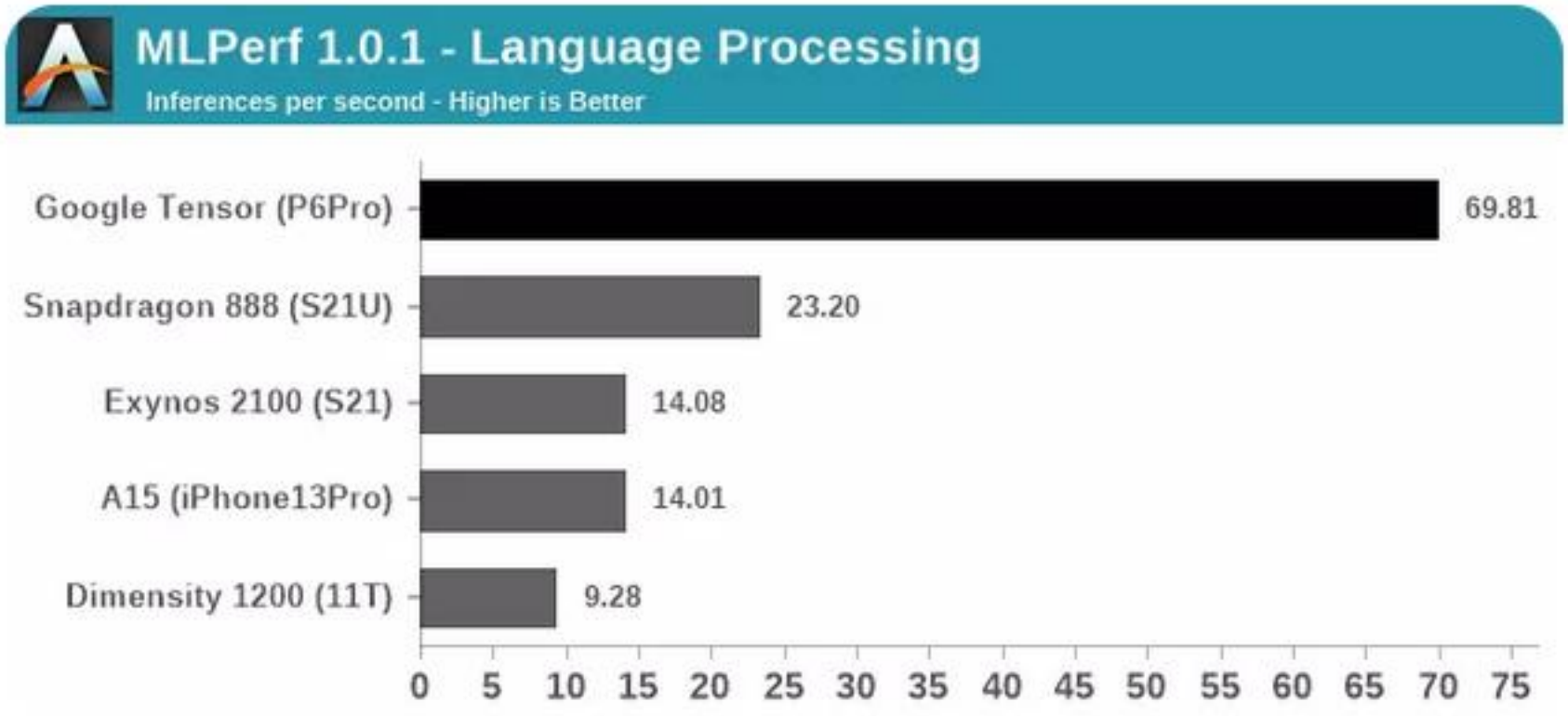
An end-to-end solution powered by machine learning



On-device centric (future)

https://www.qualcomm.com/media/documents/files/making-an-on-device-personal-assistant-a-reality.pdf

On-device processing capabilities



This completely justifies on-device processing for best user experience and privacy

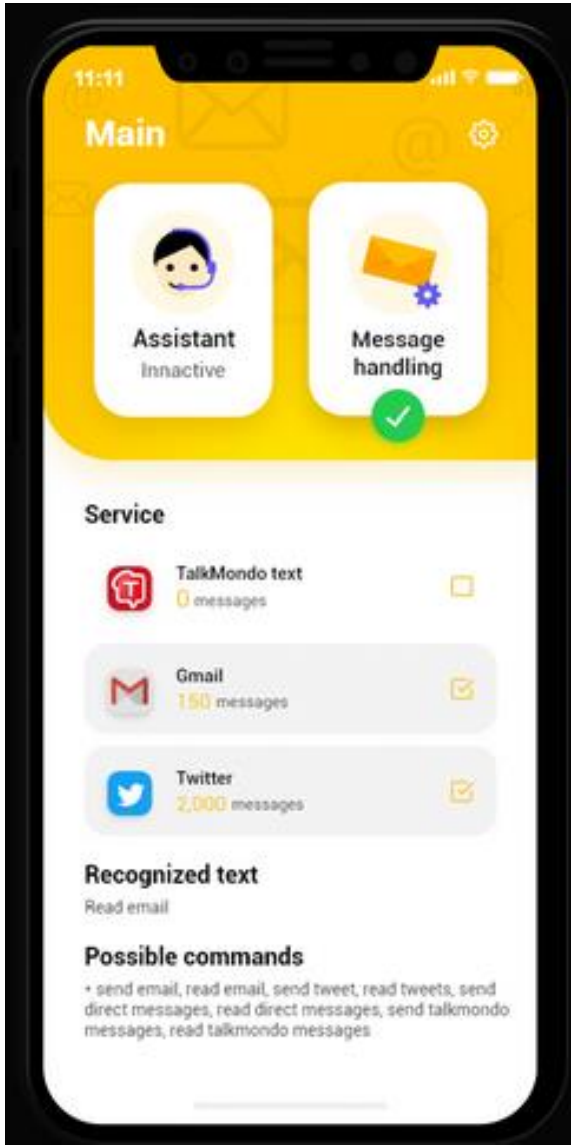
An open-source proposal from Stanford University

The Open, Privacy-Preserving Virtual Assistant

https://genie.stanford.edu/

https://www.home-assistant.io/blog/2021/12/21/stanford-genie/

REATA: your Personal Digital Assistant / Butler



- Answers telephone calls,
- Reads your emails, SMS, messages,
- Prepares text from dictation,
- Keeps track of previous conversations,
- Interogates the web for you,
- Anonymises your voice,
- Verifies your identity,
- Coaches your activities,
- Behaves like a companion / butler,
- Interacts with TalkMondo, Twitter,…
- …

Demonstration

https://www.youtube.com/watch?v=uchrARc41Gk

Conclusions and further developments

- Users will soon realise that their personal data need to be protected
- Hardware capabilities of SmartPhones are sufficient for speech and language processing
- Exploitation of voice data in the cloud in unnesserary
- Voice cloning allows for adequate anonymisation
- Adaptation of the ASR and Language models allows for improved results