Aalto University
School of Electrical
Engineering

Robust and Efficient Methods for Distributed Speech Processing

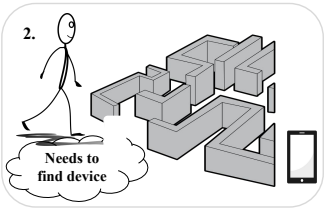
Perspectives on Coding, Enhancement and Privacy

Sneha Das

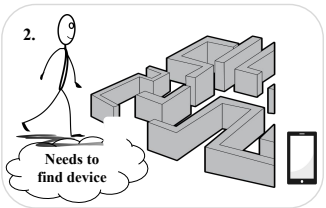
November 26, 2021

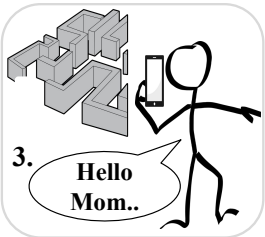
Past and present for speech communication

Past and present for speech communication



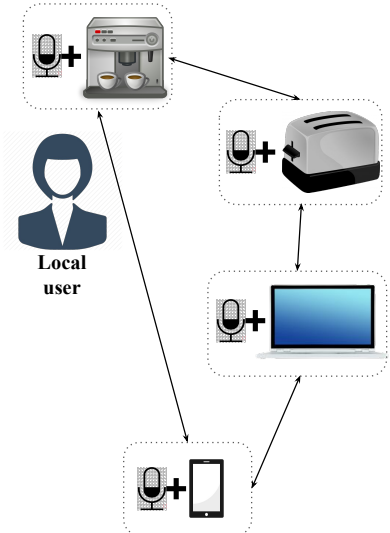
Past and present for speech communication





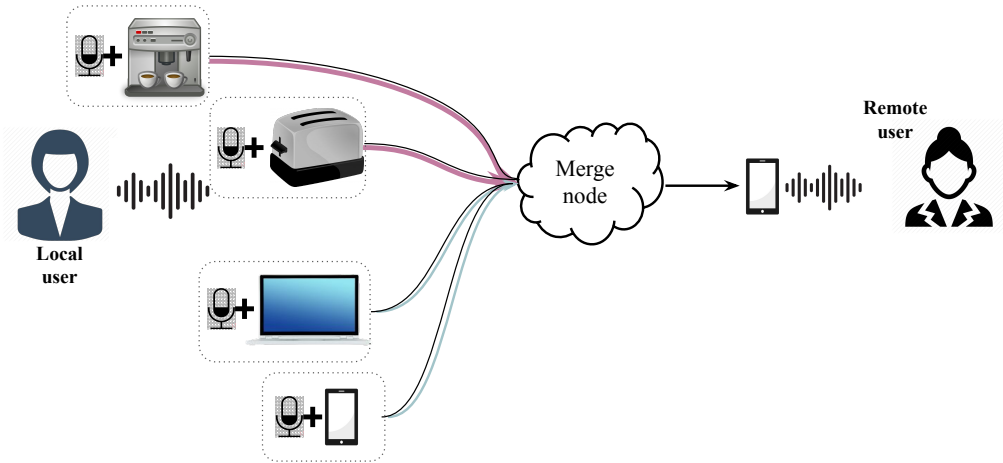
Device-centric speech communication!

Ideal: user-centric speech communication

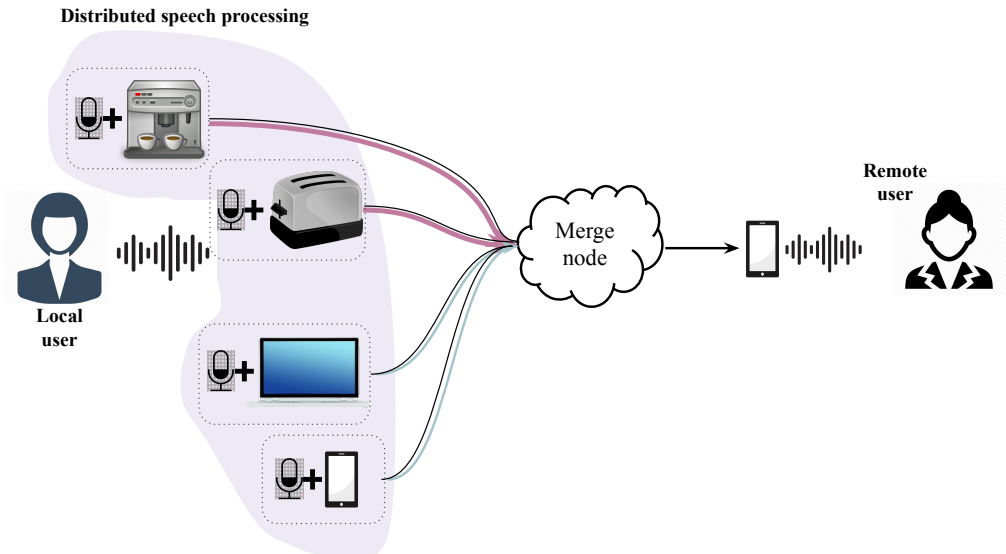


- Everyday devices are smart $\rightarrow$ embedded microphones.
- Individuals own multiple smart-devices.
- Mesh of connected devices.

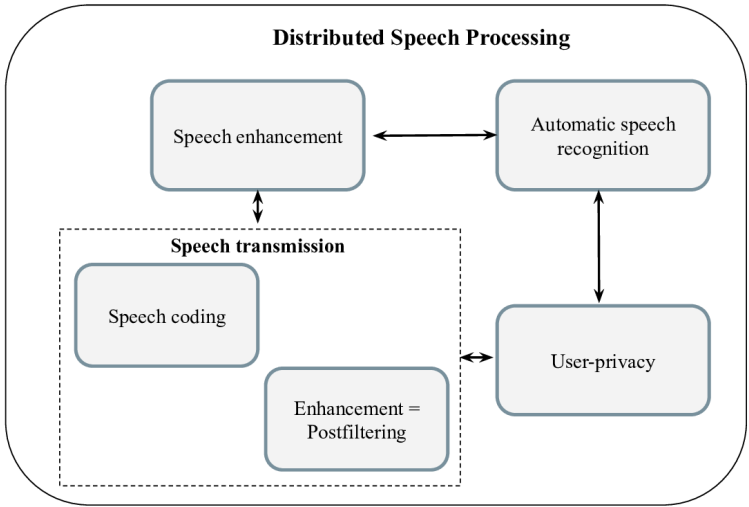
Ideal: user-centric speech communication



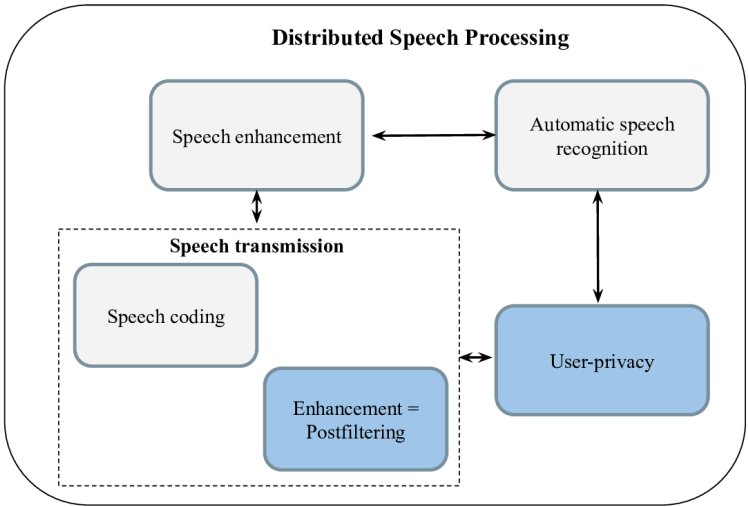
Ideal: user-centric speech communication



Distributed speech processing



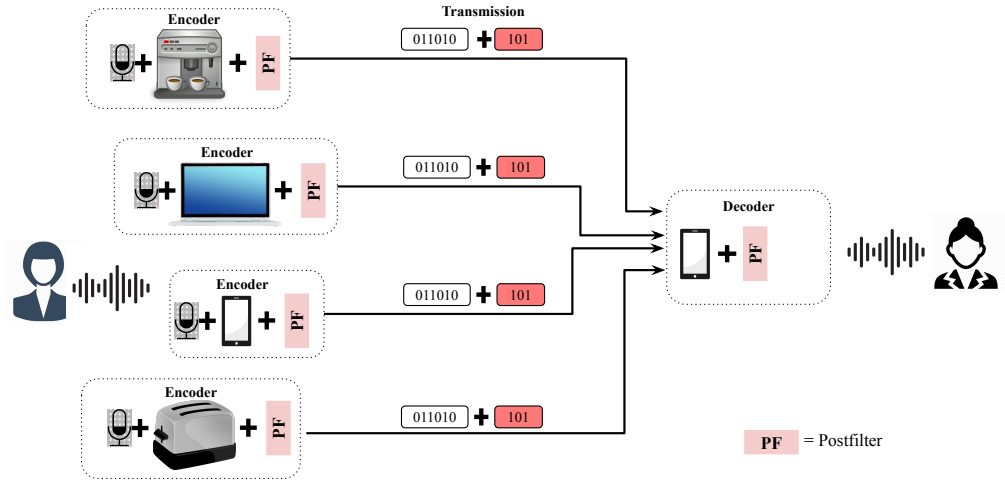
Distributed speech processing



Speech Transmission

- Speech coding $\rightarrow$ Enables speech transmission $\rightarrow$ Optimize resource consumption for transmission + transmitted speech quality.
- Postfilters $\rightarrow$ Improve signal quality at decoder.
- Conventional postfilters $\rightarrow$ (a) Processing at both encoder and decoder, (b) Additional transmitted bits, (c) Dependent on other codec functional blocks.

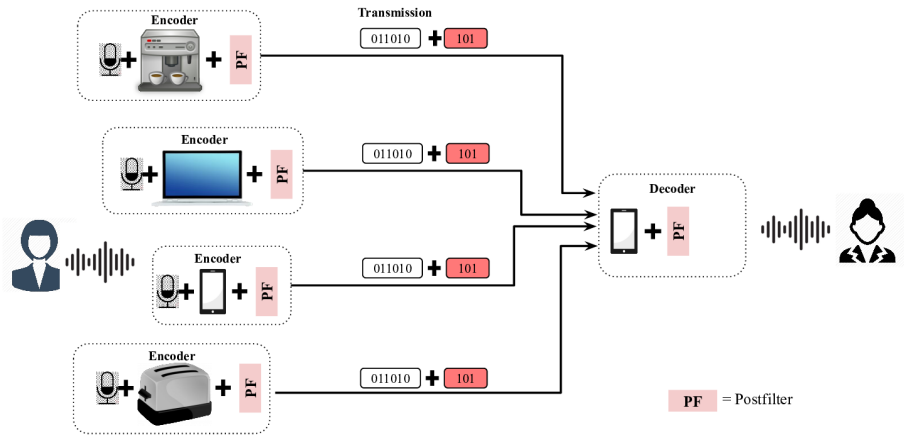
Conventional design choices



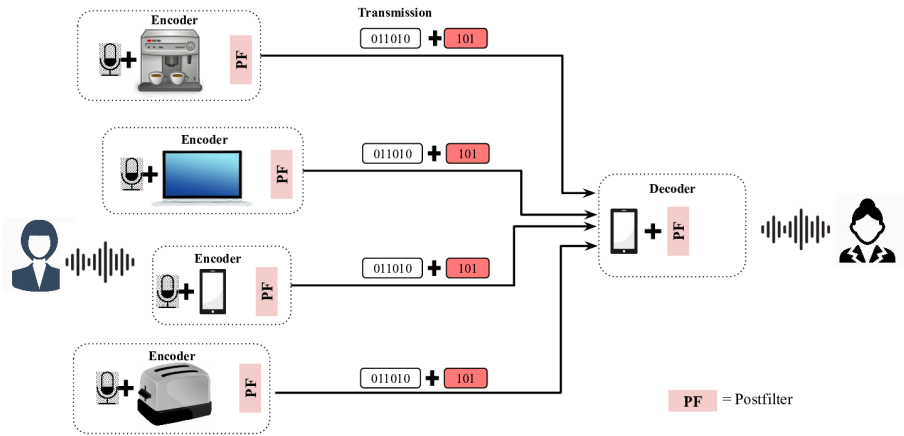
Our design choices

1. No additional bit-overhead $\rightarrow$ No additional information transmission.
2. Ensure low complexity encoder $\rightarrow$ Suitable for distributed systems.

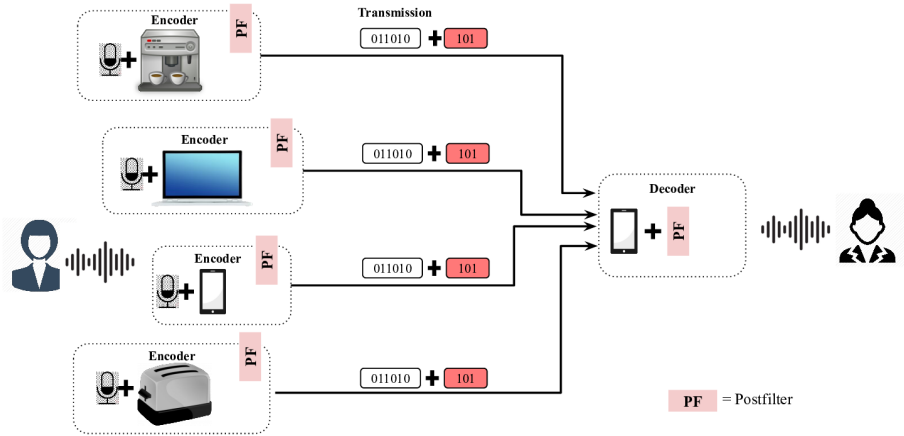
Our design choices: Decoder based postfilter



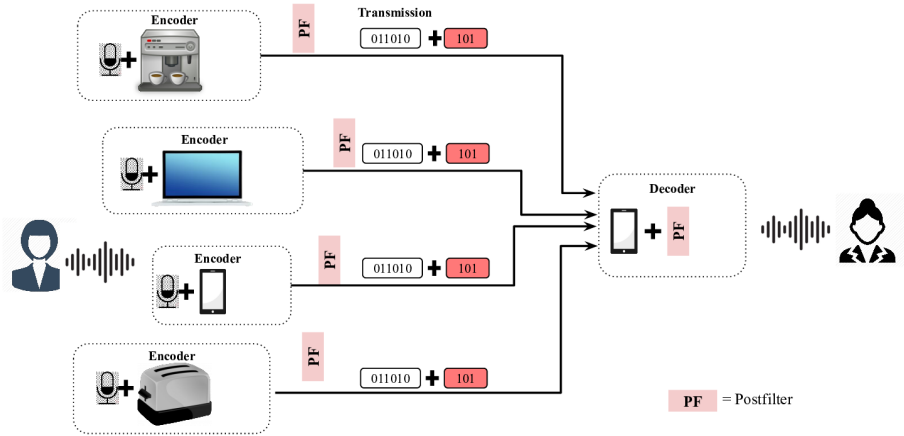
Our design choices: Decoder based postfilter



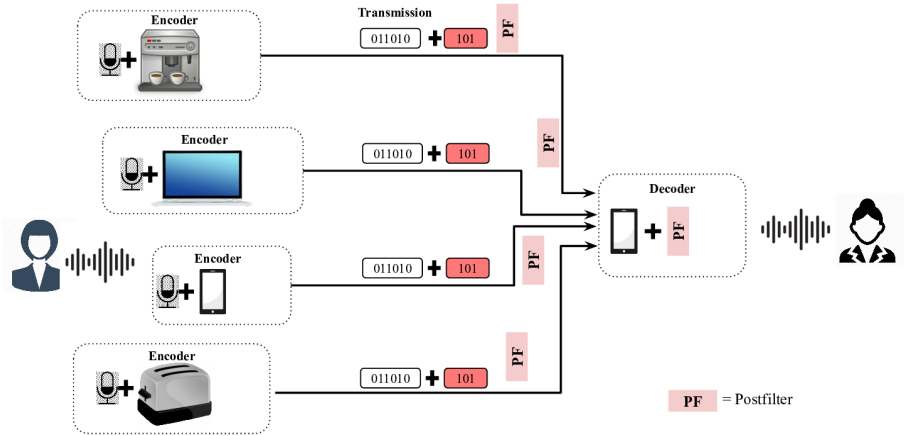
Our design choices: Decoder based postfilter



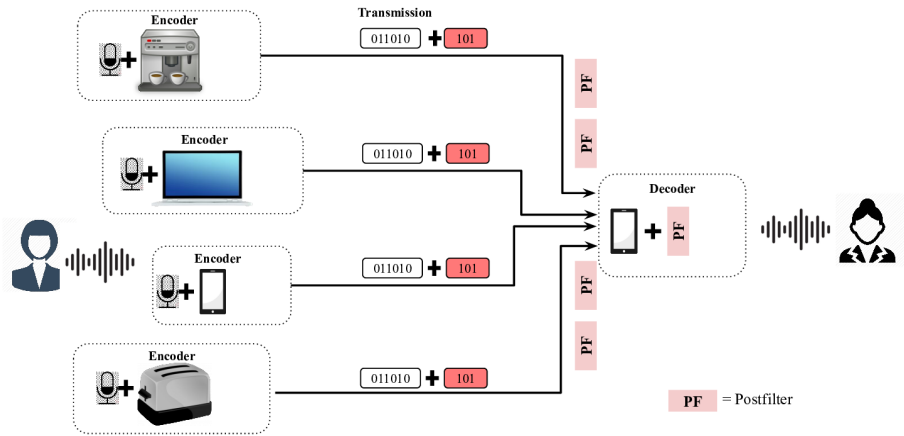
Our design choices: Decoder based postfilter



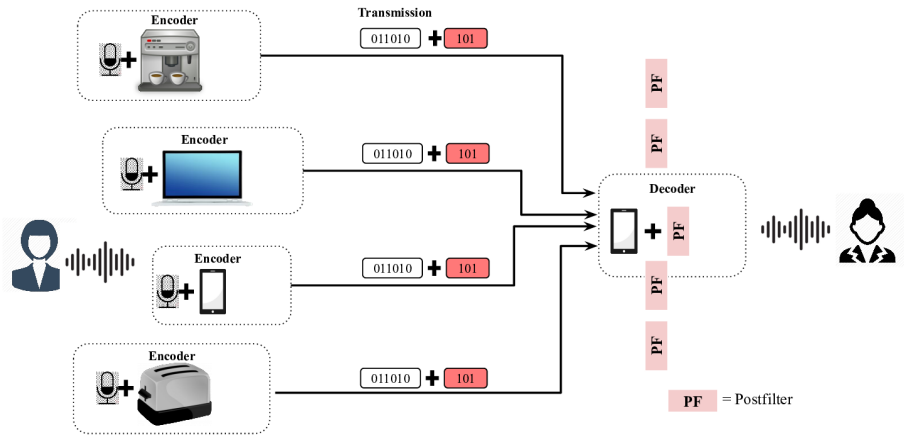
Our design choices: Decoder based postfilter



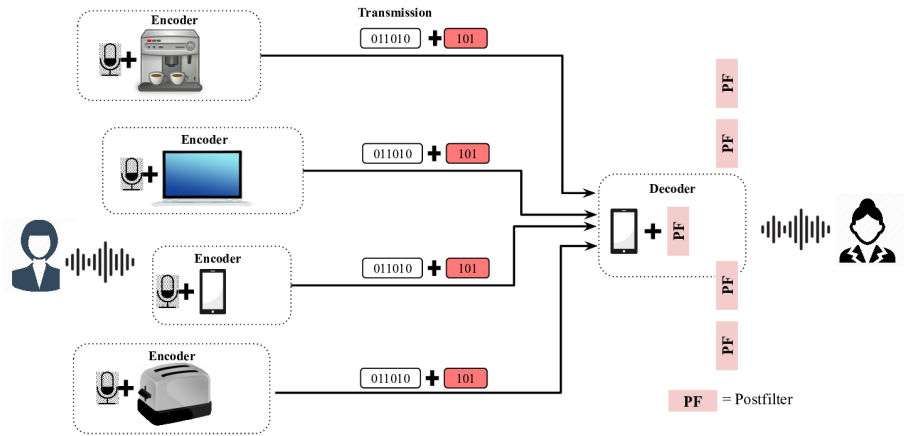
Our design choices: Decoder based postfilter



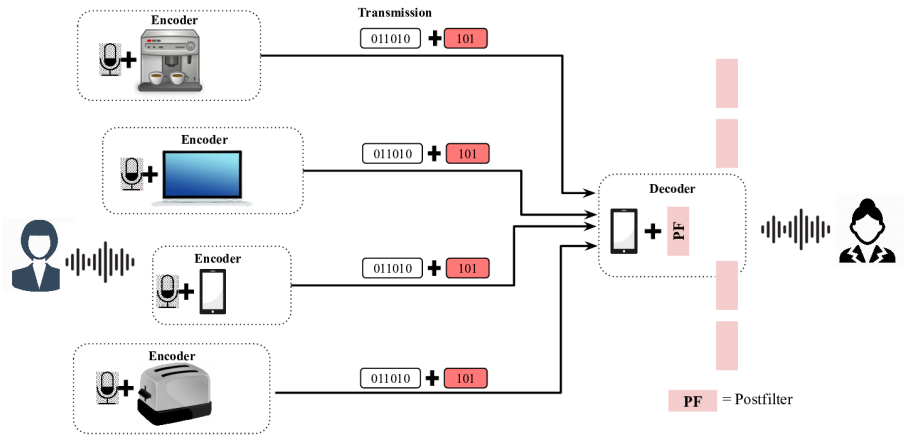
Our design choices: Decoder based postfilter



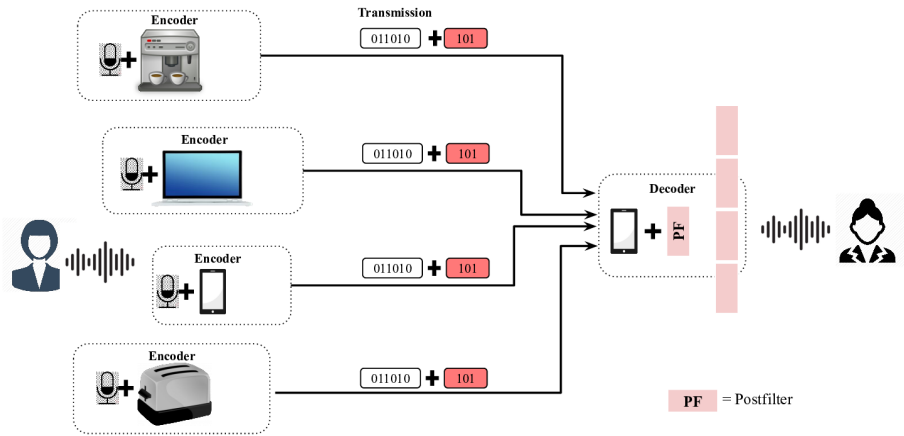
Our design choices: Decoder based postfilter



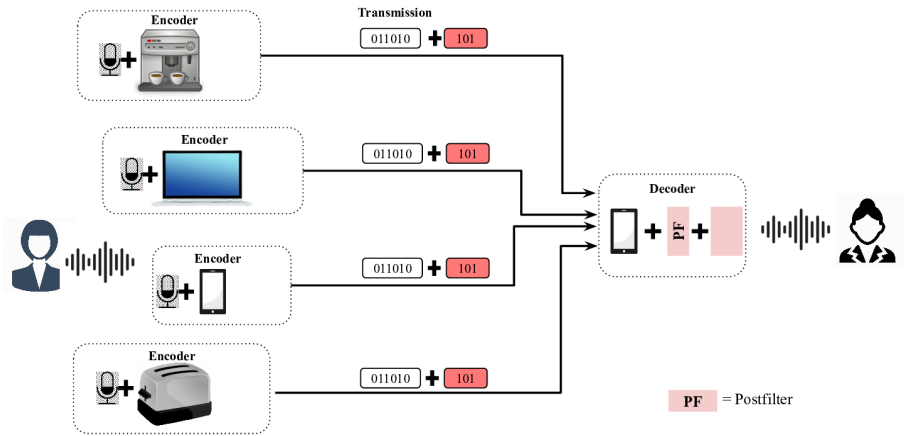
Our design choices: Decoder based postfilter



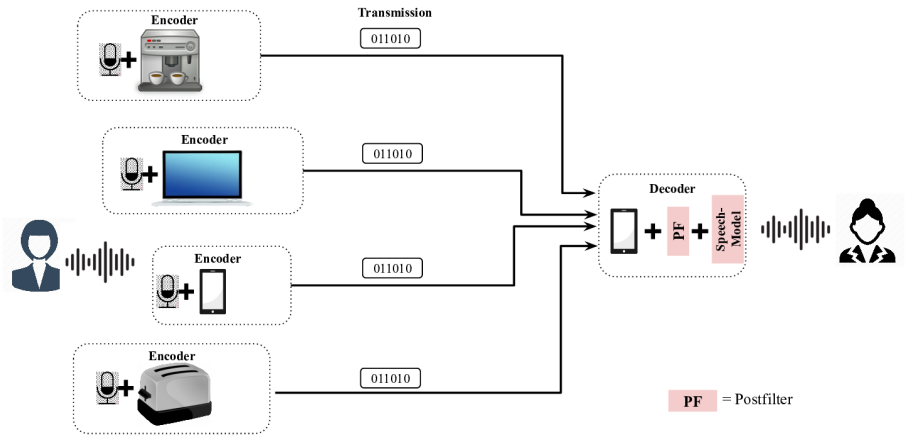
Our design choices: Decoder based postfilter

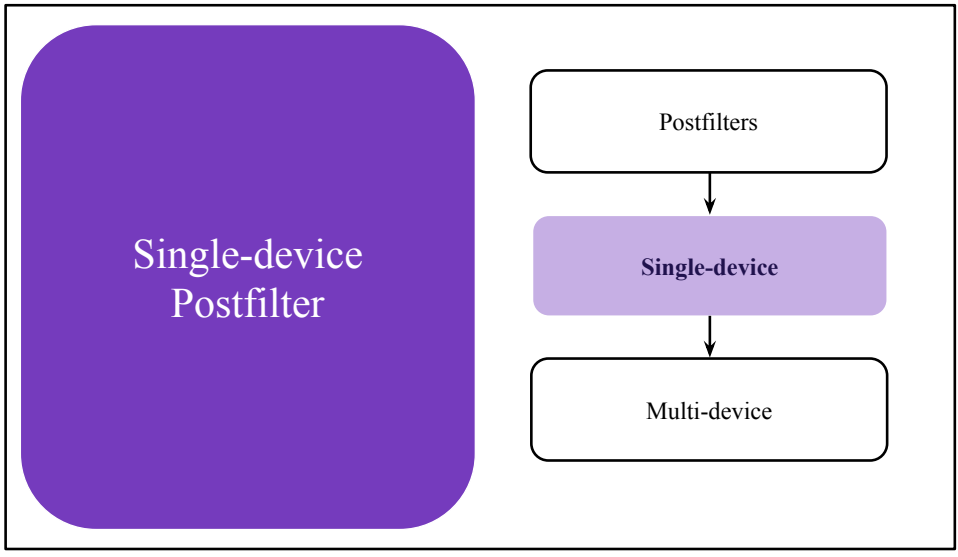


Our design choices: Decoder based postfilter



Our design choices: Decoder based postfilter

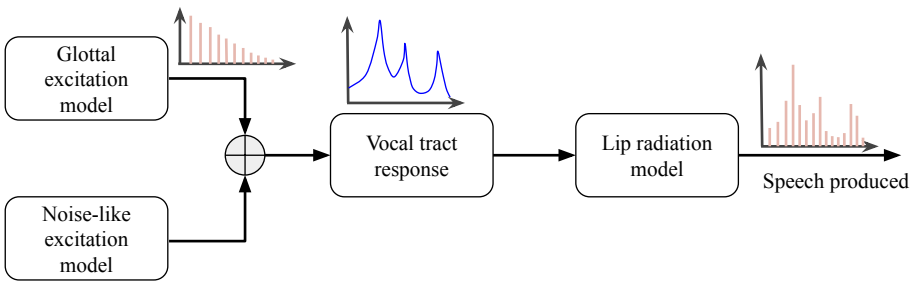



Single-device Postfilter
Postfilters
Single-device
Multi-device

Single-channel postfilter

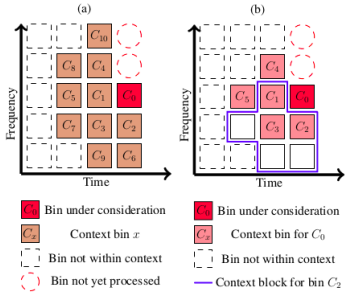
Speech production model

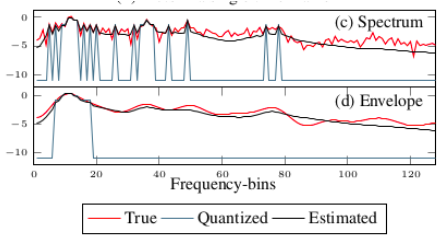
- Glottal excitation shaped by vocal tract.



Single-channel postfilter: Envelope modelling

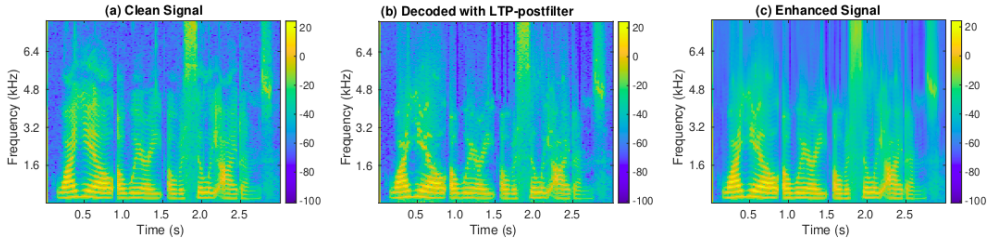
- Goal: Model time-frequency correlations in speech signals.
- Optimization: $\hat{x} = E[p(X|\mathbf{X_c} = \hat{\mathbf{x}}_c)] \quad \text{subject to} \quad l \leq X < u$

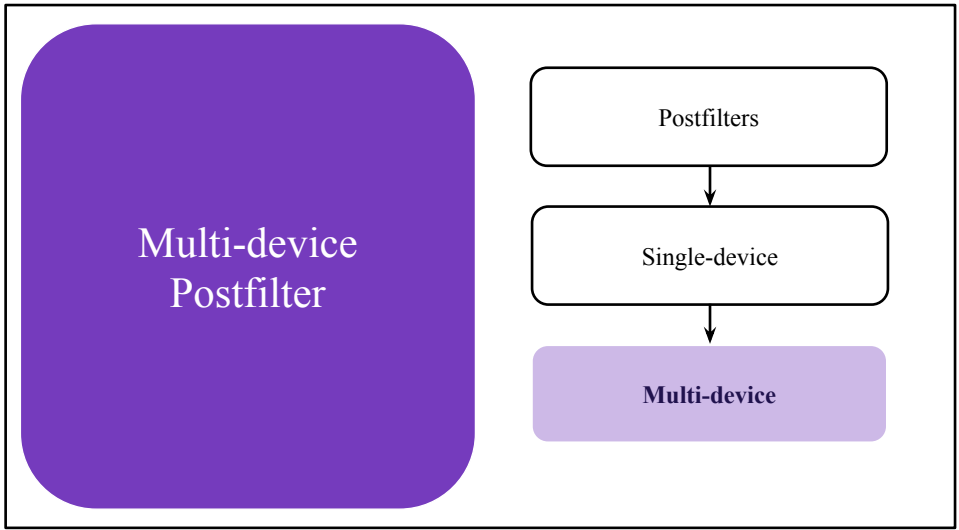




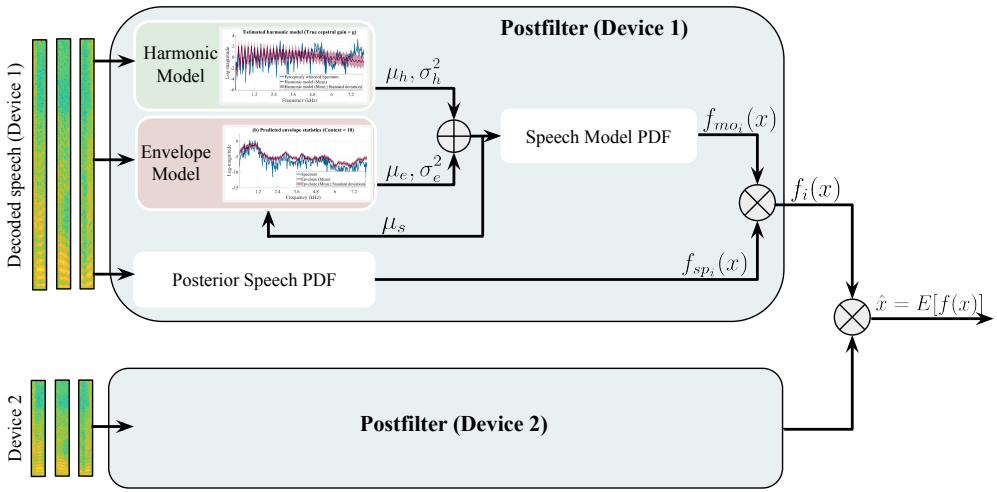
Single-channel postfilter: Harmonic modelling

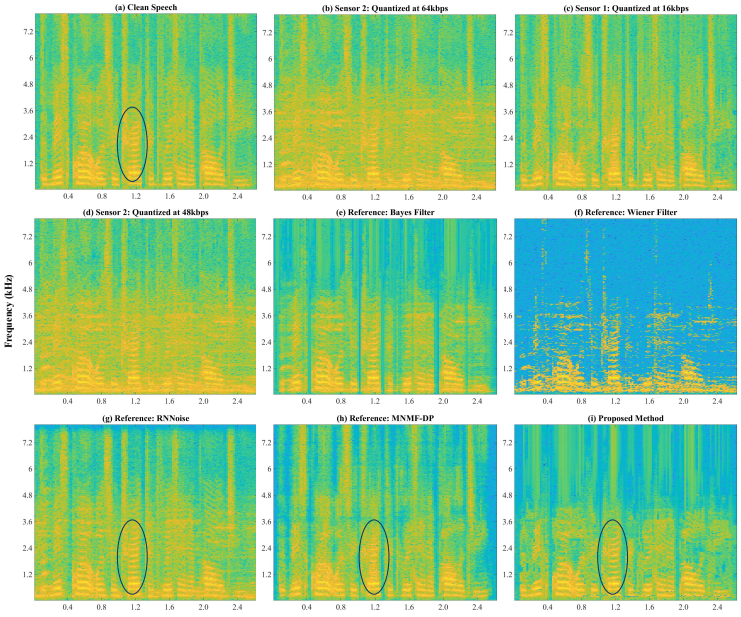
- Goal: Improve harmonic structure of the decoded signal.
- Filtering in linear domain $\rightarrow$ Multiplicative $\implies$ Additive in log-domain $\rightarrow$ $\log|\mathbf{s}| = \mathbf{x}_{\mathbf{F_0}} + \mathbf{x}_{\mathbf{env}}$; $\mathbf{x}_{\mathbf{F_0}} \rightarrow$ excitation, $\mathbf{x}_{\mathbf{env}} \rightarrow$ spectral envelope.
- Optimization: $\hat{\mathbf{s}} = \mathbf{A}^T\mathbf{d}$; MMSE to find $\mathbf{A}$ $\implies$ $\mathbf{A} = (\mathbf{D}\mathbf{D}^T)^{-1}\mathbf{D}\mathbf{S}^T$; $\mathbf{D} \rightarrow$ feature matrix.

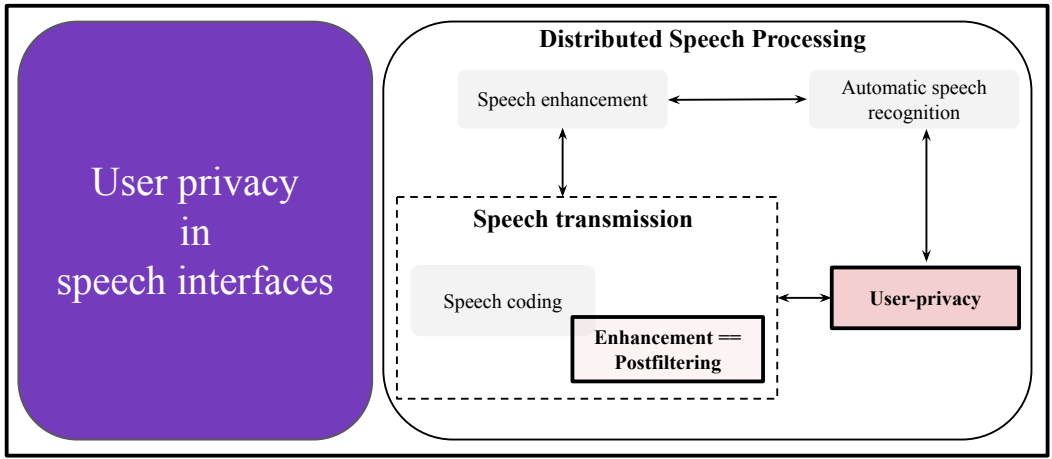



Multi-device
Postfilter
Postfilters
Single-device
Multi-device

- Goal: Merging incoming noisy and partial speech observations to obtain an enhanced representation of the speech signal.




(a) Clean Speech
(b) Sensor 2: Quantized at 64kbps
(c) Sensor 1: Quantized at 16kbps
(d) Sensor 2: Quantized at 48kbps
(e) Reference: Bayes Filter
(f) Reference: Wiener Filter
(g) Reference: RNNoise
(h) Reference: MNMF-DP
(i) Proposed Method
Frequency (kHz)
Time (s)
7.2
6
4.8
3.6
2.4
1.2
0.4
0.8
1.2
1.6
2.0
2.4


User privacy
in
speech interfaces
Distributed Speech Processing
Speech enhancement
Automatic speech recognition
Speech transmission
Speech coding
Enhancement == Postfiltering
User-privacy

Experience of Privacy

- Goal: Understanding and quantifying the perception of privacy in human communication.

Experience of Privacy

- Goal: Understanding and quantifying the perception of privacy in human communication.

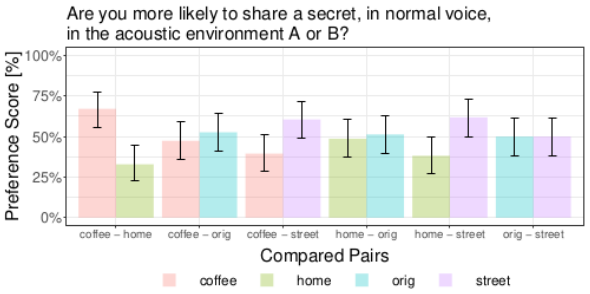
1. Speech corpus for privacy studies

- Perception of privacy in different acoustic environments.
- Supported by self-rated questionnaire.
- Two cultural settings.

2. Quantitative study

- Influence of background noise on the perception of privacy.
- Crowd-sourcing to gather responses.
- Choice models$\rightarrow$ ordering noise scenarios as per privacy.

Experience of Privacy



Insights

- Acoustic information has an influence on perception of privacy!
- Privacy preferences change based on ambient noise.
- Significant difference: coffee-shop versus home scenario.

Summary

- Enabling user-centric models of speech communication.
- Distributed speech processing, specifically for speech transmission and privacy.
  1. Make speech transmission in more robust using postfilters.
  2. Ensure efficient system in terms of complexity.
  3. First step towards understanding privacy in speech interfaces.

Thanks to all my collaborators!

Tom Bäckström
Guillaume Fuchs
Pablo Pérez Zarazaga
Anna Leschanowsky
AK Vuppala
VVV Raju...