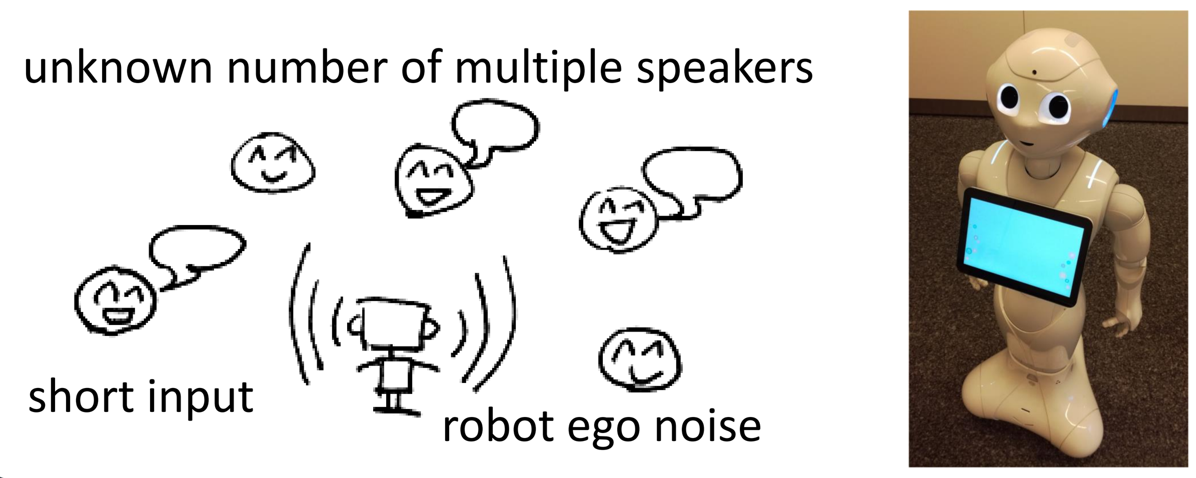
Deep Neural Networks for Multiple Speaker Detection and Localization

[1]Idiap Research Institute, Switzerland [2]EPFL, Switzerland

Weipeng He[1,2], Petr Motlicek[1], Jean-Marc Odobez[1,2]

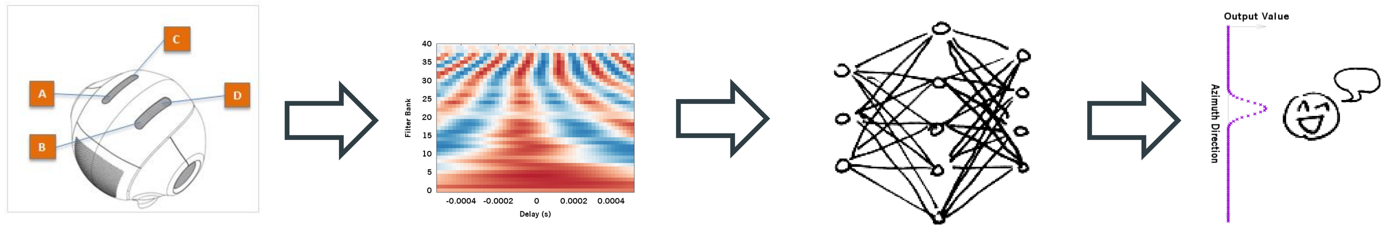
Introduction

Task: Sound source localization in real HRI scenarios



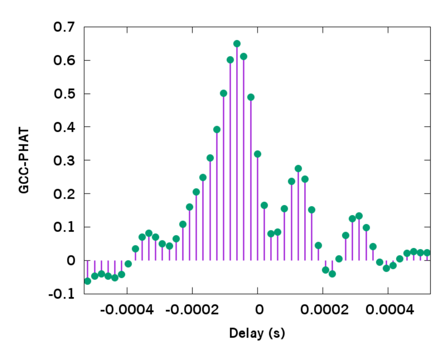
Contributions:
- Novel deep learning-based multiple sound source localization method.
- Likelihood-based output encoding handles an arbitrary number of sources.
- Investigation of three network architectures based on different motivations.
- Study sub-band cross correlation information as an input feature for better localization cues in speech mixtures.
- Collected and released a benchmark dataset of real recordings.

Approach



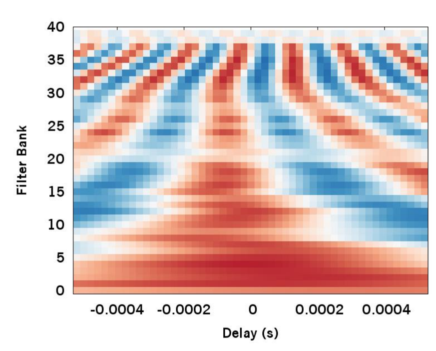
Features: Cross-correlation between pairs of microphones

GCC-PHAT coefficients



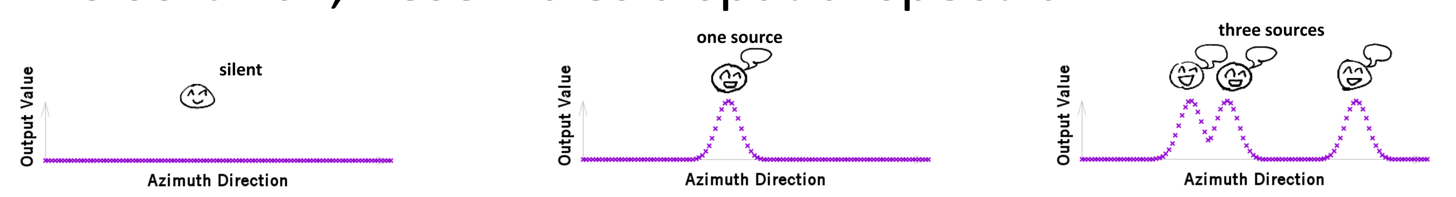
×6 pairs

GCC-PHAT on filter bank

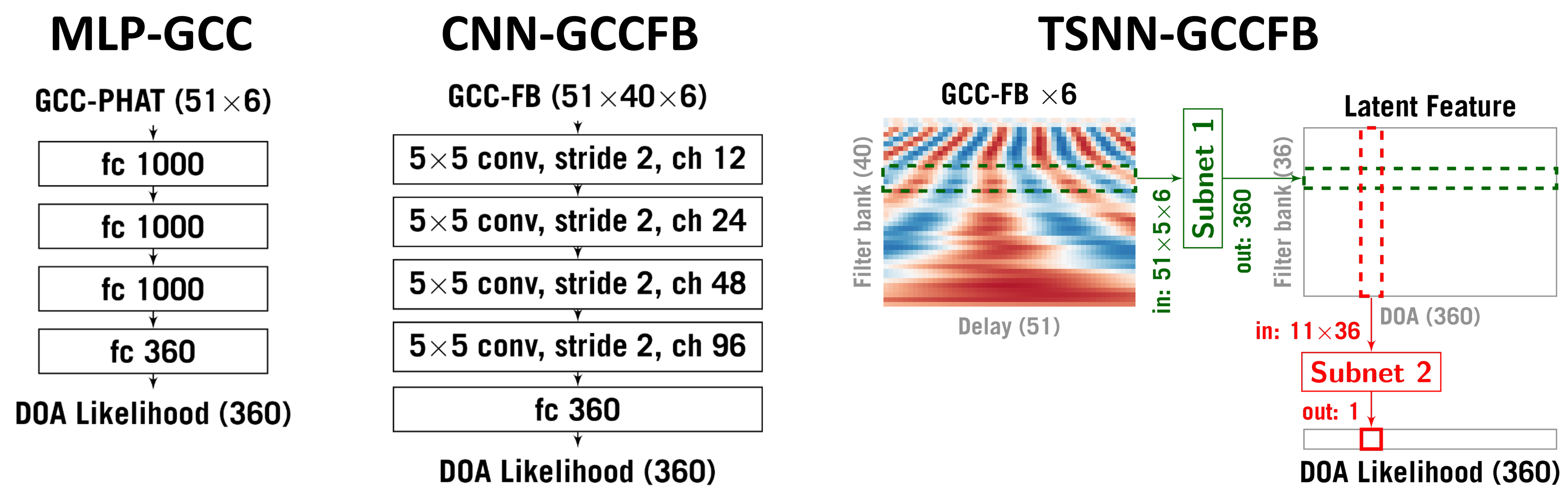


×6 pairs

Output: Likelihood of sound source being in each direction
- **Encoding:** Gaussian functions around true sources
- **Decoding:** Finding peaks
- No softmax; Resembles a spatial spectrum



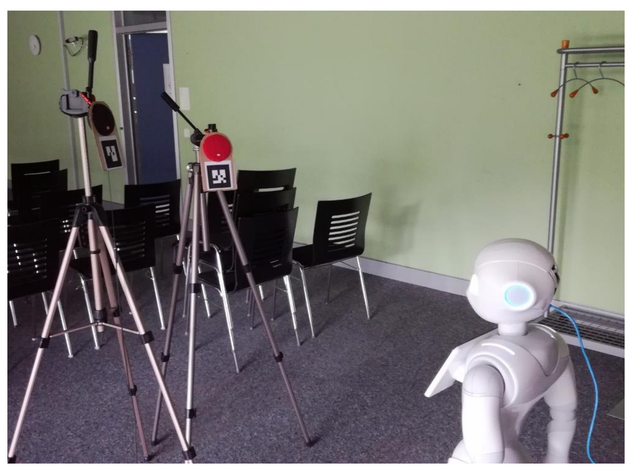
Different network architectures:
- **Multi-layer Perceptron** (MLP-GCC)
  - Basic structure (baseline)
- **Convolutional neural network** (CNN-GCCFB)
  - Convolution to reduce number of parameters
- **Two-stage network** (TSNN-GCCFB):
  - Considers the sparsity of speech signal in time-frequency points
  - First predict on sub-bands
  - Then aggregate early predictions across all frequencies.
  - Training also in 2 steps: (1) Pretrain first subnet (2) End-to-end



Experiments

Data:
- 24 hours of real recordings of Pepper.
- Up to two simultaneous speakers.

Loudspeakers
(16h train / 8h test)

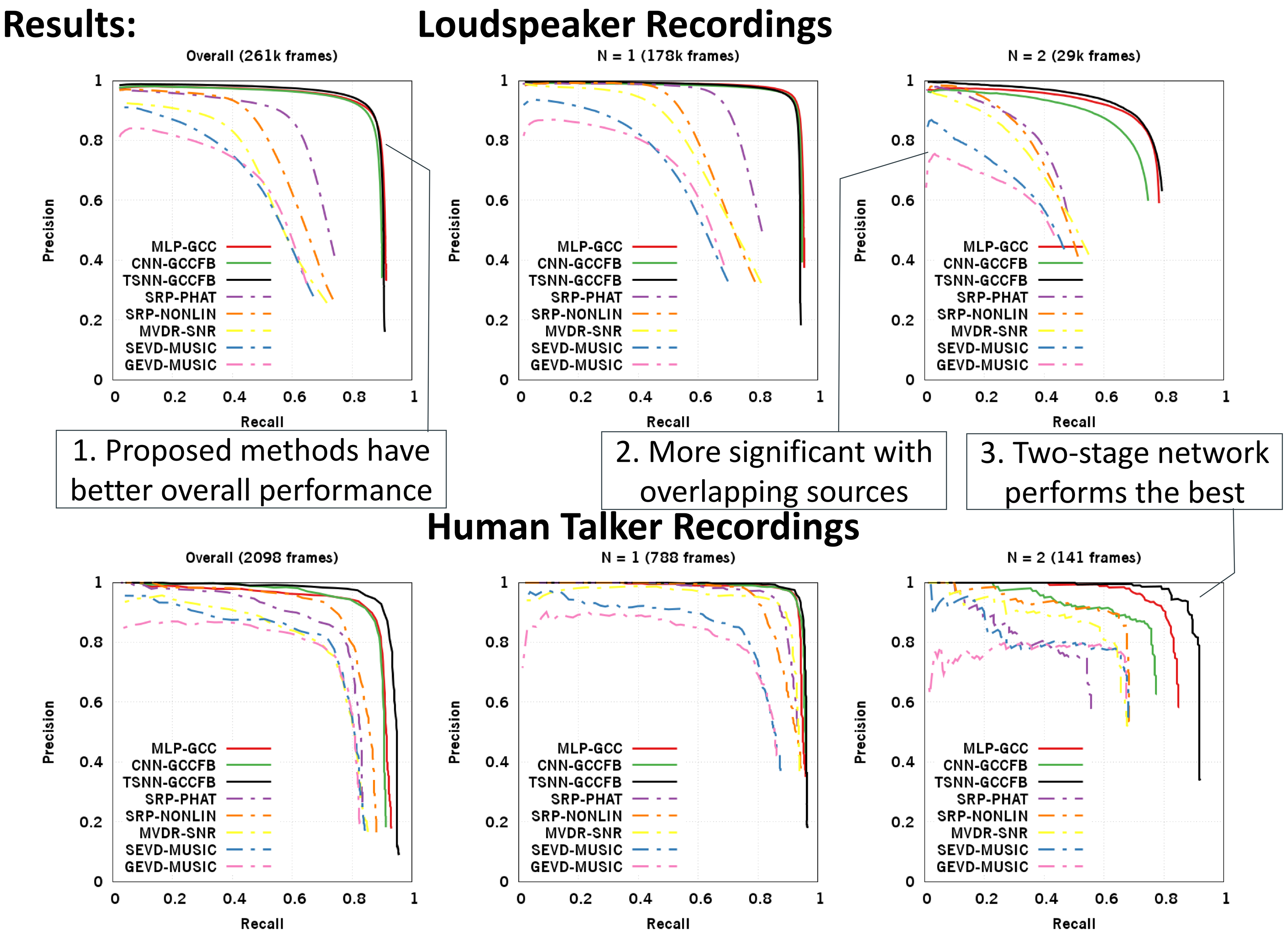


Human talkers
(4 min test)

Baseline methods:
- SRP-PHAT, MVDR, MUSIC

Evaluation:
- Number of sources is unknown => detection problem
- Prediction is correct if error < 5°
- Compute precision vs recall

Results:



Conclusion
- >90% recall and precision.
- Significantly better than popular spatial spectrum methods.

Resources
- Database: https://www.idiap.ch/dataset/sslr
- Video: https://youtu.be/_4EwuVIE_pU