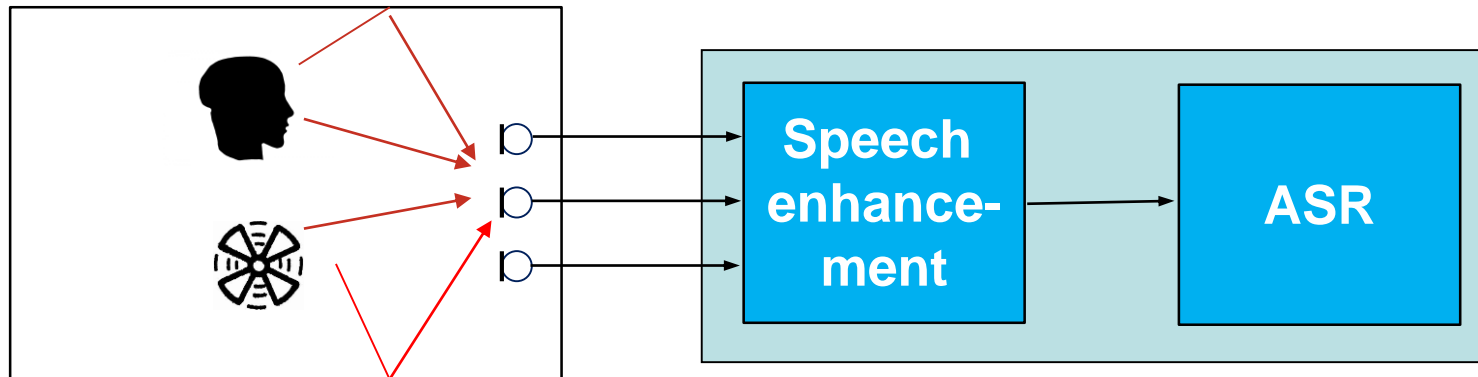


# **Part VII. Summary**

**Reinhold Haeb-Umbach &  
Tomohiro Nakatani**

---

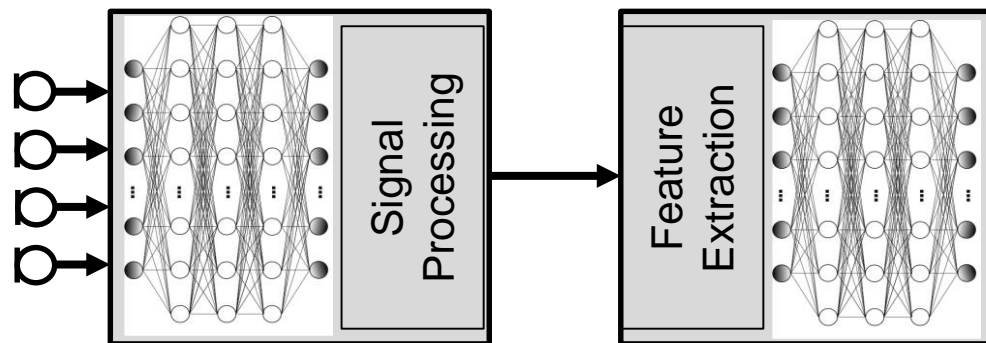
# Combination of speech enhancement and ASR



- Speech enhancement for ASR is recommended
  - If phase (spatial) information present in multi-channel input can be exploited, which would be lost in traditional ASR feature representations
    - Acoustic beamforming
  - If distortions exist, which introduce huge variability in frame-based ASR processing
    - Reverberation
    - Multiple concurrent speakers
  - Where excellent signal processing solutions exist (which can be further improved by deep learning)
    - MIMO acoustic echo cancellation (not treated in this tutorial)

# Speech enhancement by DSP and DNN

- We have seen many examples in this tutorial of combinations of traditional signal processing and deep learning techniques
- Compared to pure DSP they offer several advantages
  - Leverage training data
  - Overcome restrictions of simplifying modeling assumptions otherwise necessary to obtain tractable solutions
- Compared to pure DNN they offer the following advantages
  - Less data hungry
  - Better interpretable
  - Can adapt to test data via unsupervised learning



# Trends

- End-to-end trained (enhancement + ASR) systems
- DNNs will gain ever more grounds
  - Future DNNs may include microphone array functionality
  - Compact DNN on device
- Multimodal processing
  - Vision, bio sensors, brain activities, etc.

# Future challenges

- Get rid of simplifying assumptions
  - E.g., #speakers constant and known in a mixture
  - Transcribe realistic meeting scenarios
- Leverage huge amounts of unlabeled speech and audio
  - From supervised learning to unsupervised learning enabled by signal processing
- Cope with more challenging environments / applications
  - E.g., CHiME-5 dinner party transcription (WER > 40%)
- Lack of domain/environment specific training data
  - „Speech processing in the wild“

Fortunately, there is still a lot to be done!

Get started<sup>1</sup>, and enjoy working in this fascinating field!



Tutorial  
preparation

<sup>1</sup> Get hands-on experience using the various pointers to software found in this tutorial!