Connected Visual Reality – High Quality Audio Visual Communication in Heterogeneous Networks

RWTH AACHEN UNIVERSITY

ind Institute of Communication Systems and Data Processing
Prof. Dr.-Ing. Peter Vary

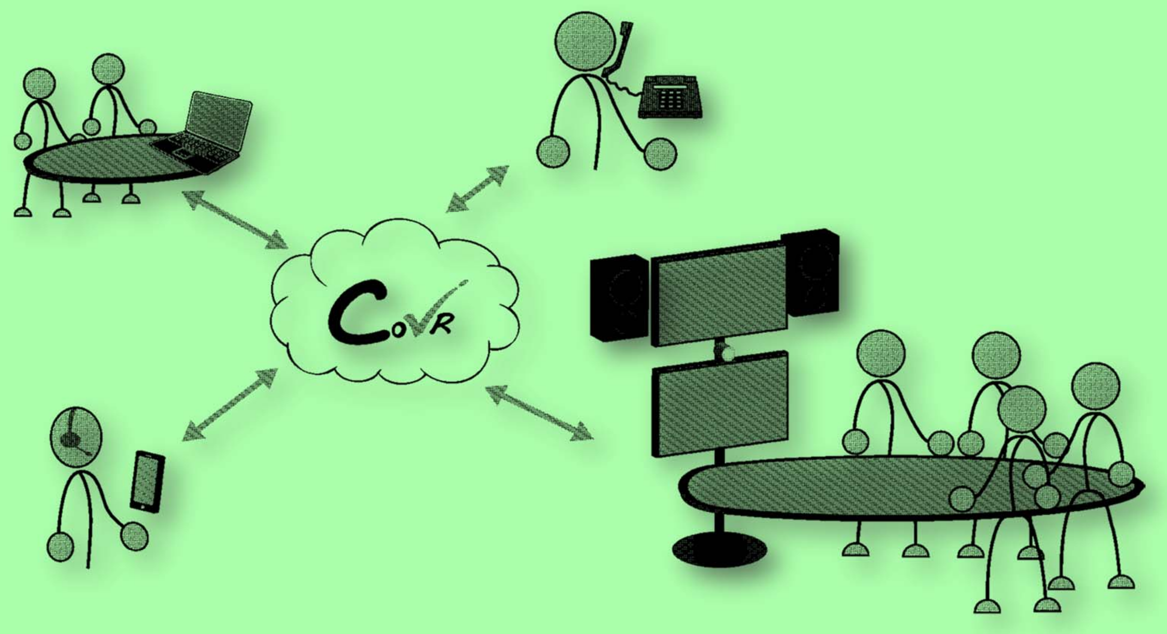
Magnus Schäfer, Florian Heese,
Thomas Schlien, Christiane Antweiler

Audio Signal Enhancement

Project Goals - Audio

- Signal processing and enhancement with multiple microphones
- Source localization
- Homogenization of speech quality in heterogeneous networks
- Spatial rendering and scene composition

Scenario



System Design

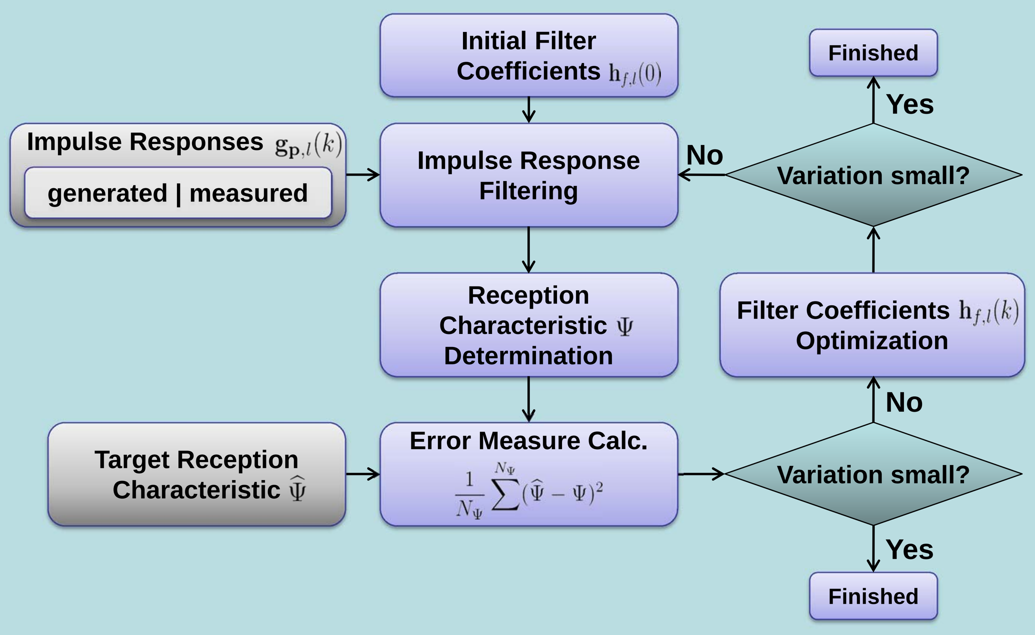
- Exchange of information between audio and video analysis
- Transmission of audio signals and meta data
- Central processing and mixing

Microphone Array Processing

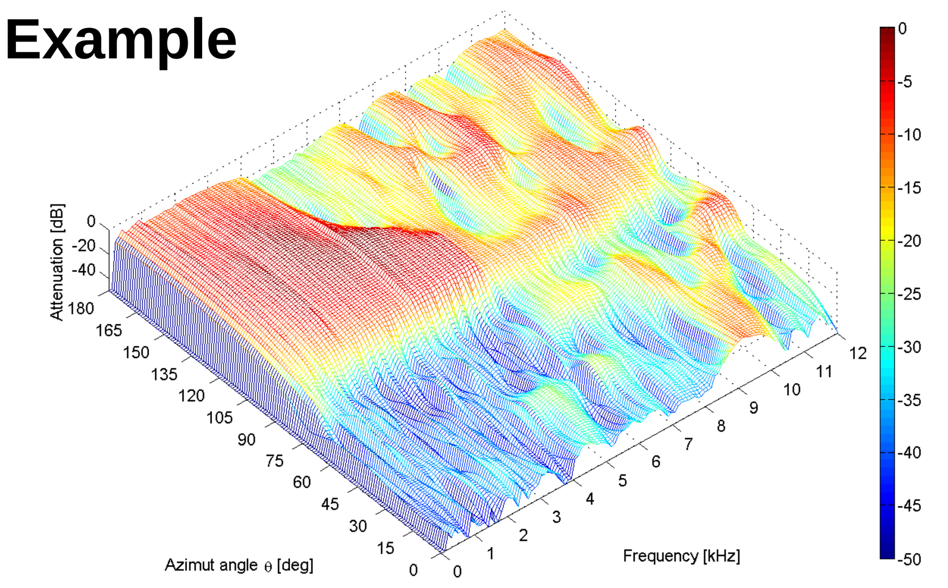
Motivation

- Reverberation and level of background noise sufficiently low in conference rooms
- Spatial separation is normally possible in conferencing scenarios
  → Microphone array: efficient way to amplify desired region while attenuating other regions (e.g. noise)

Filter Coefficient Determination [1]



Example



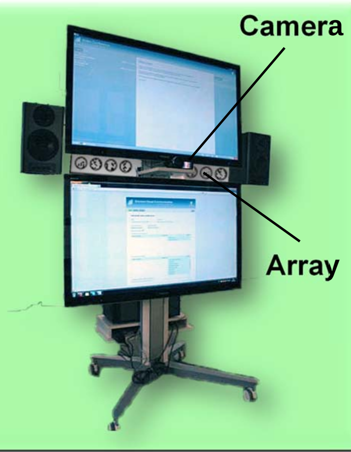
Based on measured impulse responses in anechoic room
- Filter length: 48
- Microphone distances: 3 - 3 - 3 - 30 - 3 - 3 - 3 cm
- Spatial aliasing frequency: approx. 5700 Hz

Interaction Audio and Video

- Face detection by video processing returns coordinates
- Extract speakers audio (using precomputed filter sets)
- Speaker VAD
- Determine most active speaker(s)
  - Region of interest video encoding
  - Audio rendering

DEMO:

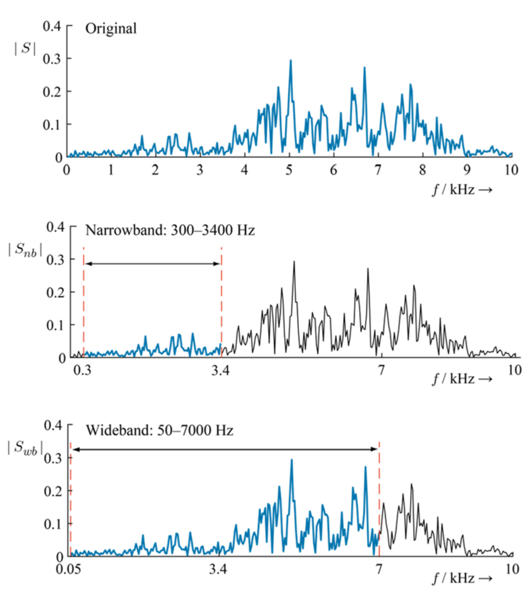
Separate two zones at far end using array of 8 microphones and the proposed beamformer



Artificial Bandwidth Extension (BWE)

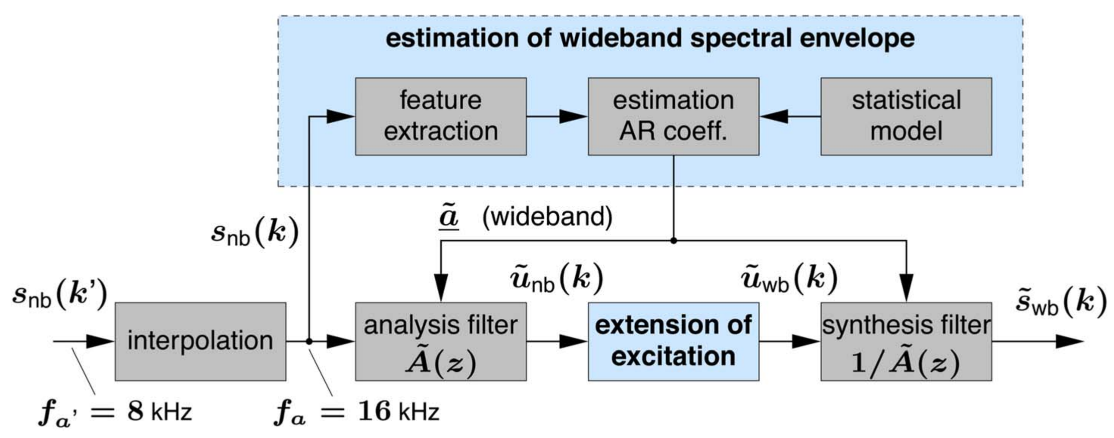
Motivation

- **Limited frequency range of "telephone speech"**
  - Degraded quality and intelligibility
- **Wideband speech coding**
  - Changes in network required
- **BWE of narrowband speech**
  - No modification of infrastructure
  - Improvement of subjective speech quality & intelligibility



Spectrum of a sharp /s/

Bandwidth Extension Algorithm [2]



Separate extension of

- Spectral envelope (AR coefficients $\tilde{\underline{a}}$)
- Excitation signal $\tilde{u}(k)$

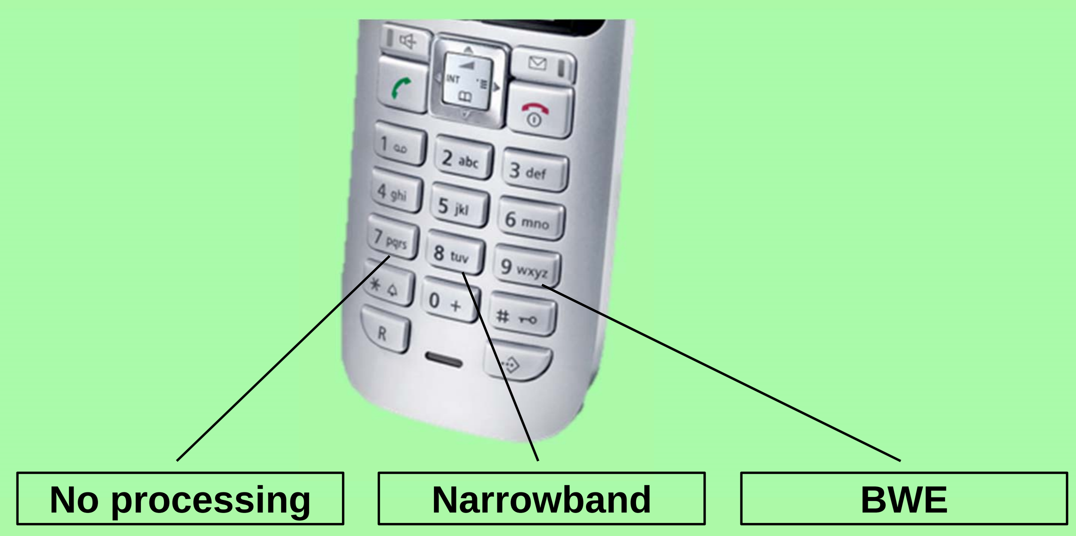
Transparent with respect to the narrowband signal

VoIP-Server: Freeswitch

FreeSWITCH

- Open-source multi platform Software-PBX (GPL)
- Interfaces:
  - VoIP (SIP, IAX, H.323, …)
  - Traditional telephone network (PSTN, analog and digital)
- Many supported codecs such as G.722, G.719 codec (wideband / fullband)
- Extend Freeswitch with new modules
  - Audio signal processing API

DEMO: Real-time BWE conference dial-in

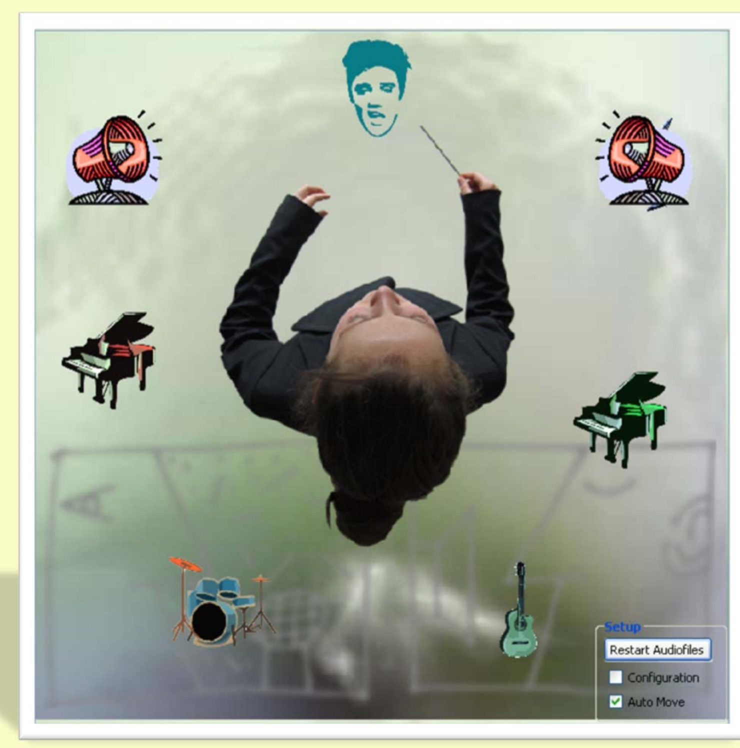


Spatial Rendering and Mixing

Motivation

- **Innovative rendering & mixing in central processing unit**
- **Meta-data driven concept**
  - Idea: Transmission of several single audio channels + meta-data
- **Head related impulse responses (HRIRs)**
  - Continuous-azimuth acquisition of HRIRs
  - NLMS driven adaptation algorithm
  - Perfect sweeps as excitation signal
- **Real-time demonstration based on RTProc [3]**

DEMO: Binaural rendering with continuous HRTFs [4, 5]



Acknowledgment

This work was co-funded by the German federal state North Rhine Westphalia (NRW) and the European Union (European Regional Development Fund: Investing In Your Future).

[1] M. Schäfer , F. Heese, J. Wernerus, and P. Vary, "Numerical Near Field Optimization of Weighted Delay-and-Sum Microphone Arrays", International Workshop on Acoustic Signal Enhancement (IWAENC), Aachen, 2012.

[2] P. Jax and P. Vary, "On artificial bandwidth extension of telephone speech," Signal Processing, vol. 83, no. 8, pp.1707–1719, Aug. 2003.

[3] H. Krüger and P. Vary, "RTPROC: A System for Rapid Real-Time Prototyping in Audio Signal Processing," Proceedings of IEEE/ACM International Symposium on Distributed Simulation and Real Time Applications, pp. 311–314, Vancouver, 2008.

[4] G. Enzner, "Analysis and Optimal Control of LMS-Type Adaptive Filtering for Continuous-Azimuth Acquisition of Head Related Impulse Responses," Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), Las Vegas, 2008.

[5] C. Antweiler and G. Enzner, "Perfect Sequence LMS for Rapid Acquisition of Continuous-Azimuth Head Related Impulse Responses," in IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, 2009.