Aalto University
School of Electrical
Engineering

Directional Audio Coding - A Perception-Based Method for Spatial Sound Reproduction

Mikko-Ville Laitinen

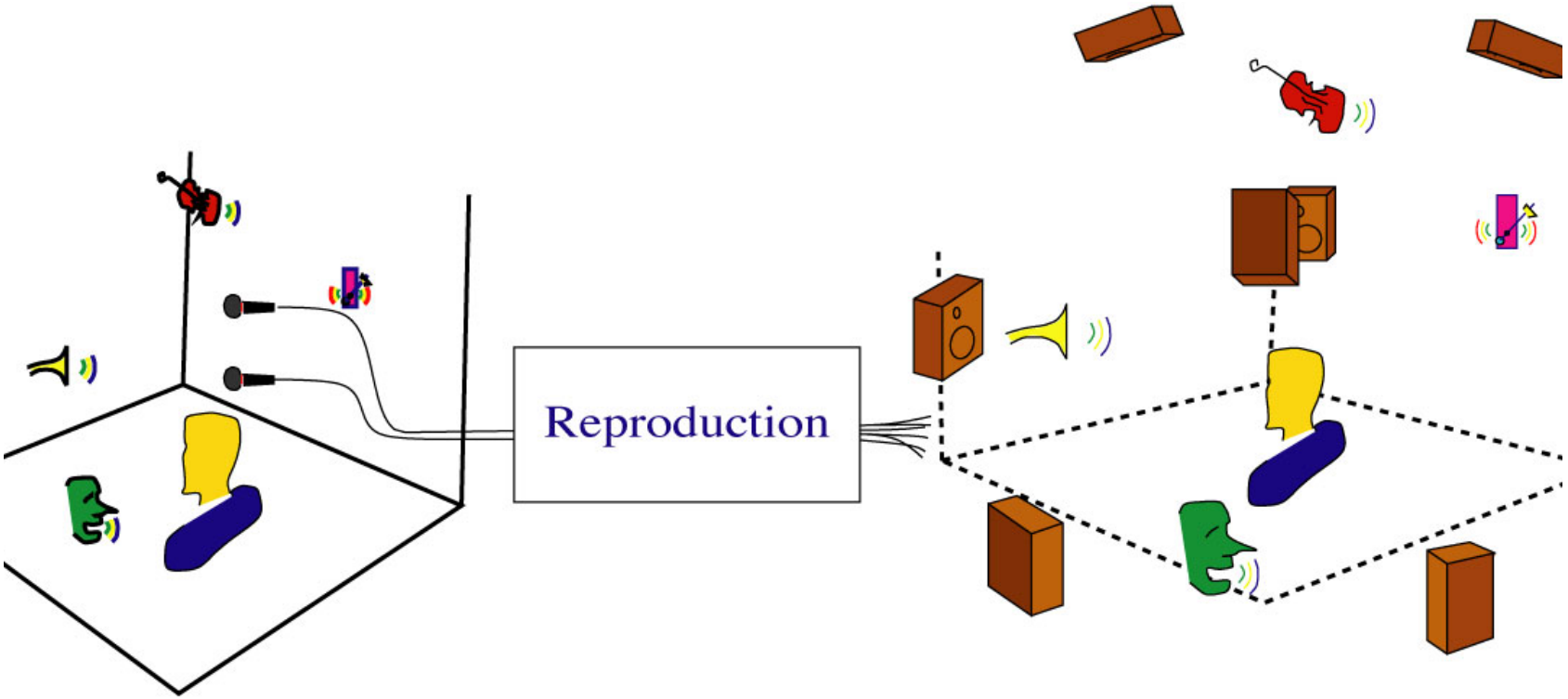
Department of Signal Processing and
Acoustics, Aalto University, Espoo, Finland

Oct 24-26, 2012
2nd SPLab Workshop

Outline

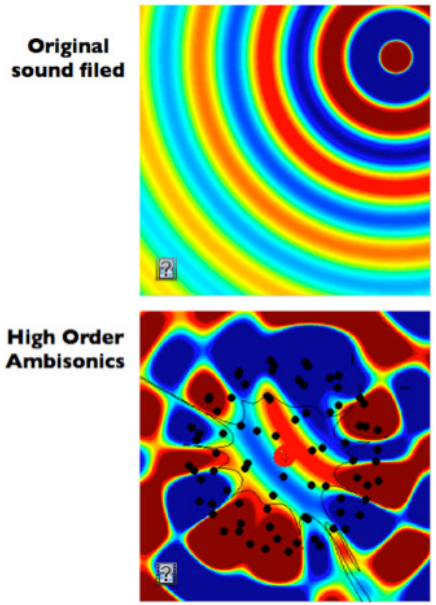
- Introduction
- Background
- Directional Audio Coding (DirAC)
- Applications
- Evaluation
- Summary

Introduction



Background
Possible approaches

- Reproduce spatial sound in a way that
  - The physical sound field is accurate
  - The human perception is accurate



Background
Sound field reconstruction methods

- Methods
  - Wave field synthesis
  - (Higher-order) Ambisonics
- Reproduced sound field is compared to original sound field
  - Perfect reconstruction -> perfect perceptual quality!
  - What is the perceptual impact of possible deviations?

Background
Perception-based methods

- Methods
  - DirAC
  - MPEG Surround
  - Binaural cue coding (BCC)
- Original sound field is not aimed to be reconstructed
- Analyze features from the sound field that are important to spatial sound perception
- Reproduce these features accurately
  - The aim is obtain equal *perception* of the reproduced sound field as the original sound field

Directional audio coding (DirAC)
Assumptions about spatial hearing

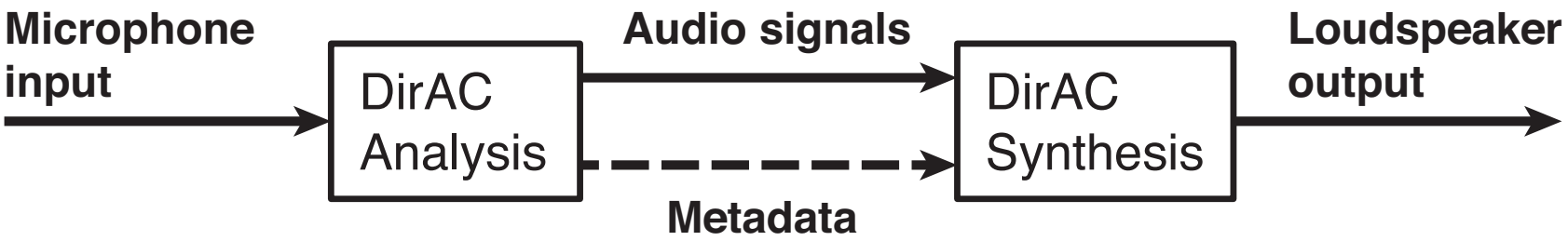
- Human hearing analyzes sound in frequency bands
  - Equivalent rectangular bandwidth (ERB)
- At each frequency band can be perceived
  - Interaural time difference (ITD)
  - Interaural level difference (ILD)
  - Interaural coherence
- If two sinusoids are at the same frequency band, they cannot be localized individually

Directional audio coding (DirAC)
Basic idea

- Analyze features of the sound field that translate into these parameters in hearing:

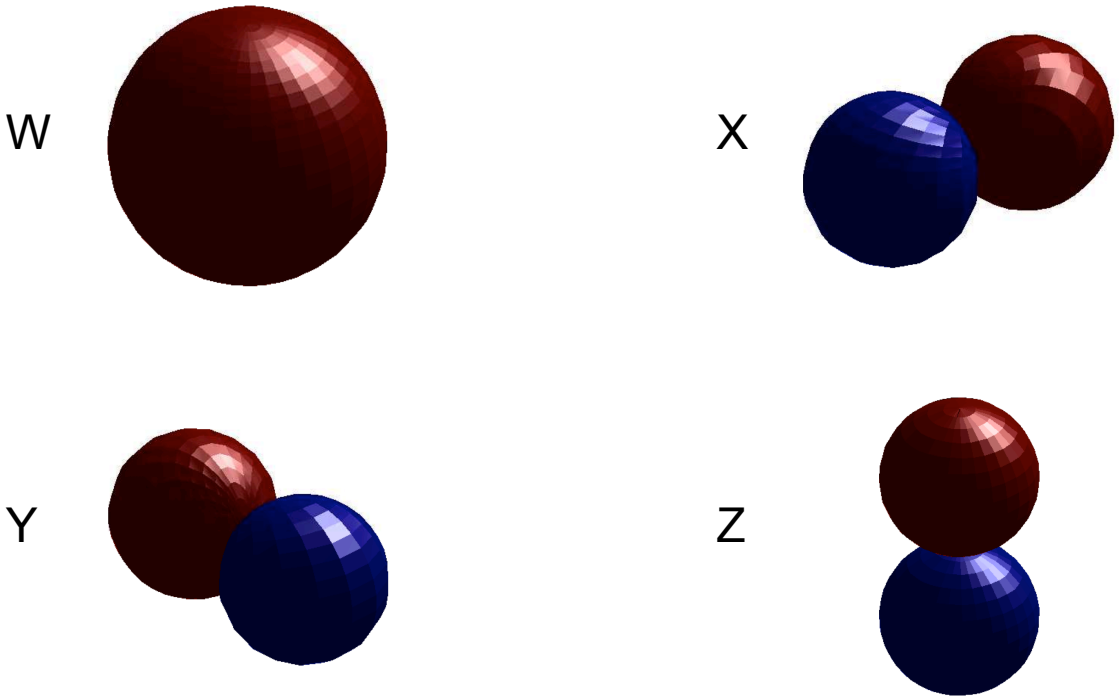
Sound field		Human hearing
Direction of arrival (DOA)	→	ITD, ILD
Diffuseness	→	Coherence

Directional audio coding (DirAC)
Basic block diagram



Directional audio coding (DirAC)
Microphone input

- B-format



Directional audio coding (DirAC)
Pressure and velocity approximates

- $W \sim p$
- $[X\mathbf{e}_x \; Y\mathbf{e}_y \; Z\mathbf{e}_z] \sim \mathbf{u}$

Directional audio coding (DirAC)
Energetic analysis of sound field

- Intensity

$$\mathbf{I} = p\mathbf{u}$$

  - pressure x velocity

- Energy density

$$E = \frac{1}{2}\rho_0 \left(\frac{p^2}{Z_0^2} + ||\mathbf{u}||^2 \right)$$

  - pressure squared + velocity squared

Directional audio coding (DirAC)
DirAC analysis

- Analyzed separately for each frequency band:
  - Direction of arrival

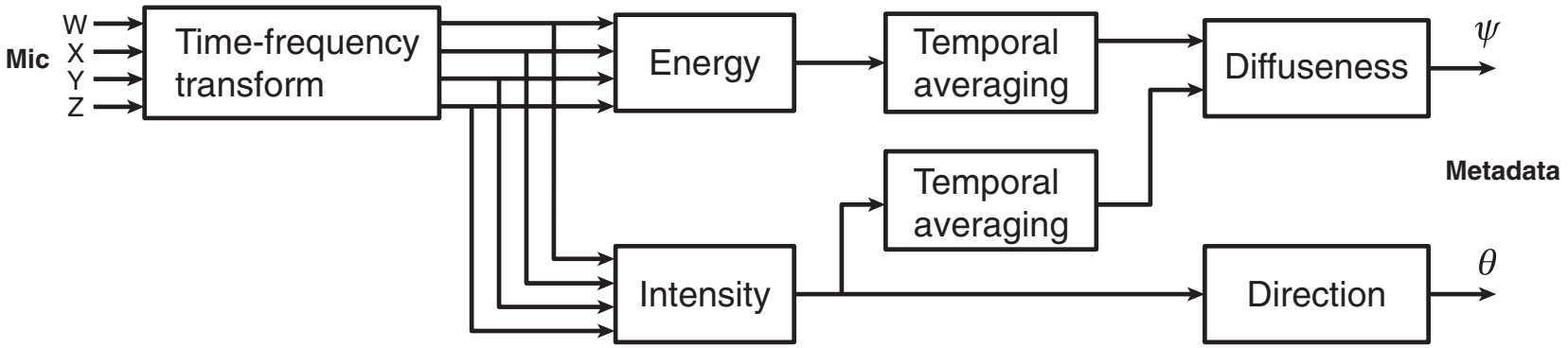
$$\theta = -\mathbf{I}$$

  - Diffuseness

$$\psi = 1 - \frac{\|\mathrm{E}\{\mathbf{I}\}\|}{c\mathrm{E}\{E\}}$$

(ratio of propagating and total energies)

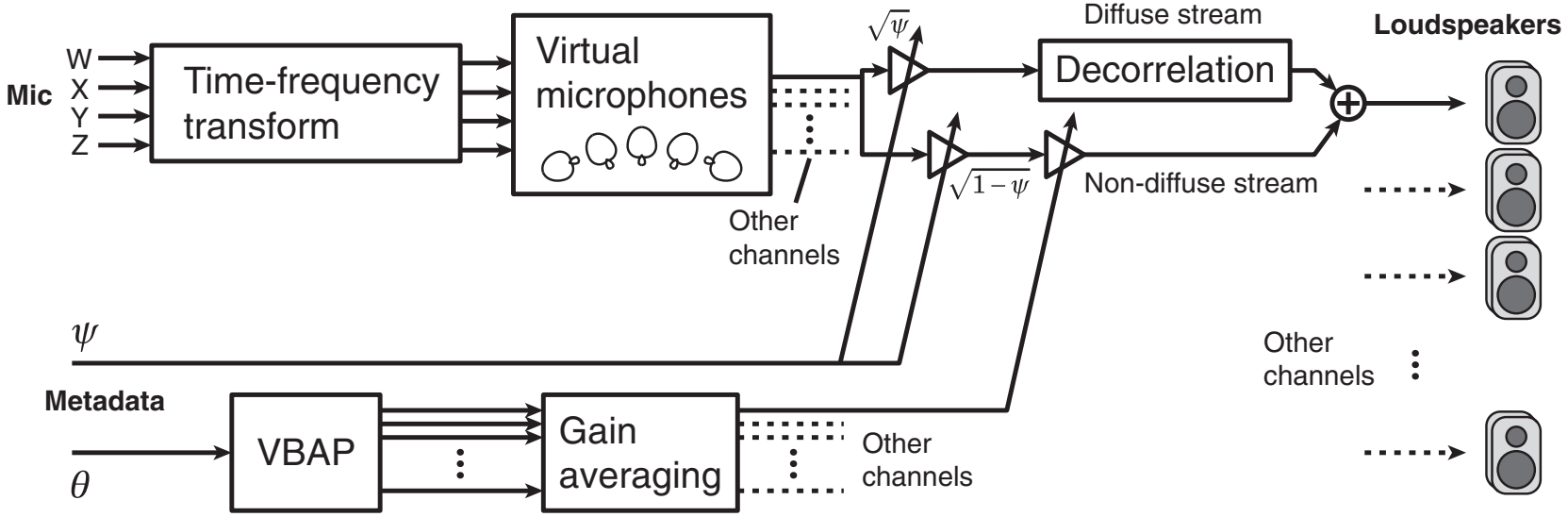
Directional audio coding (DirAC)
DirAC analysis



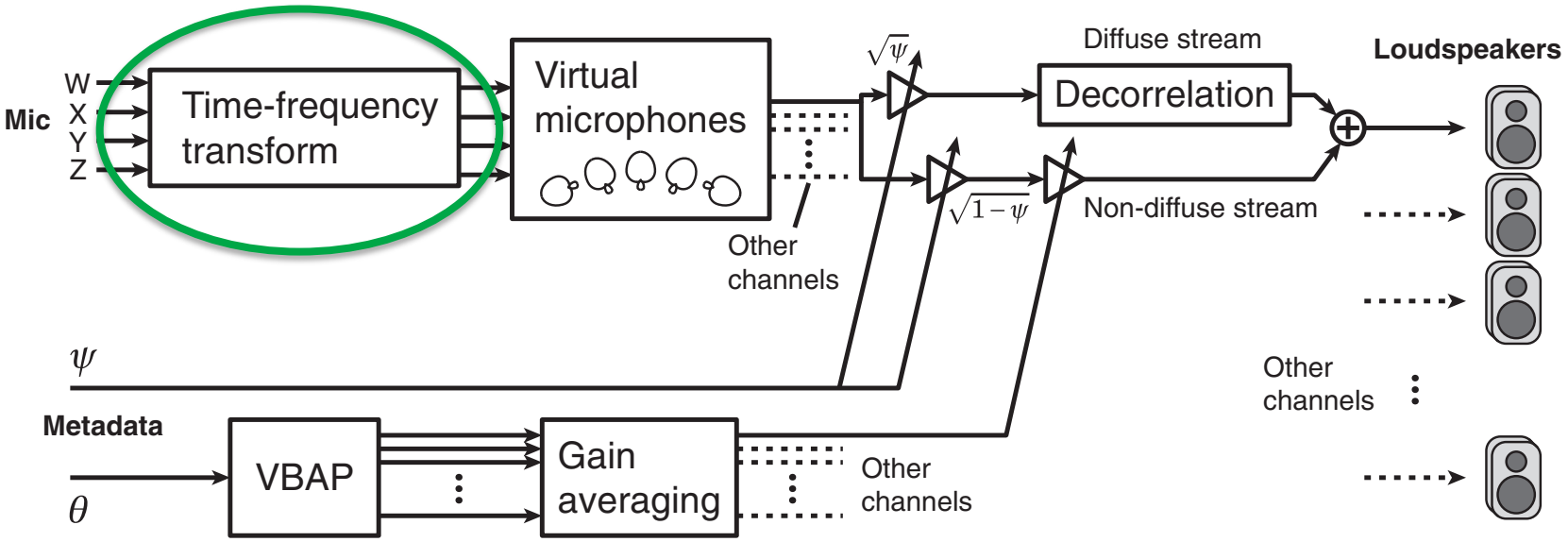
Directional audio coding (DirAC)
Time-frequency transform

- Short-time Fourier transform (STFT)
- Multi-resolution STFT
- Filterbank

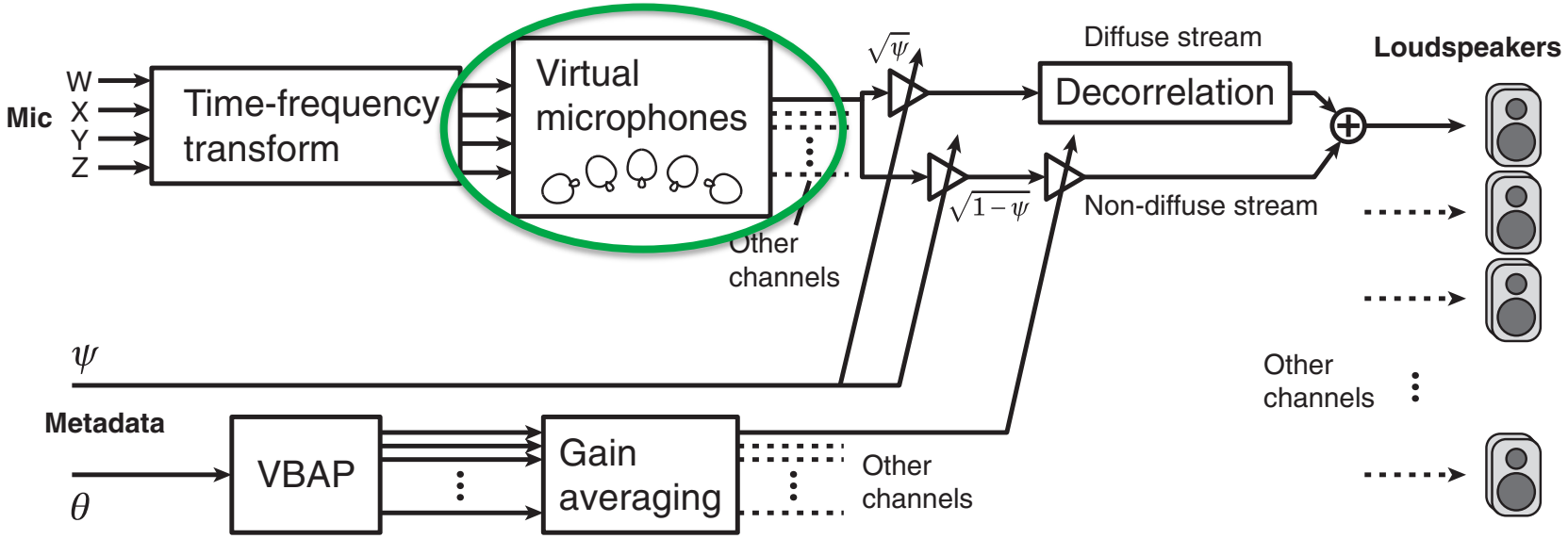
Directional audio coding (DirAC)
DirAC synthesis



Directional audio coding (DirAC)
DirAC synthesis

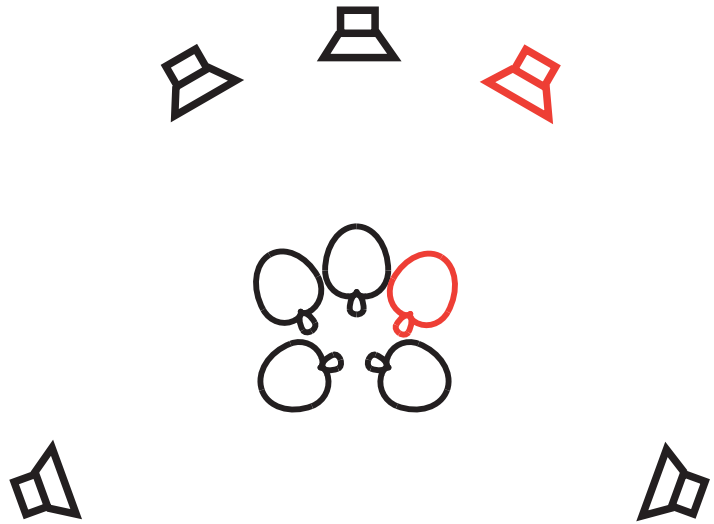


Directional audio coding (DirAC)
DirAC synthesis

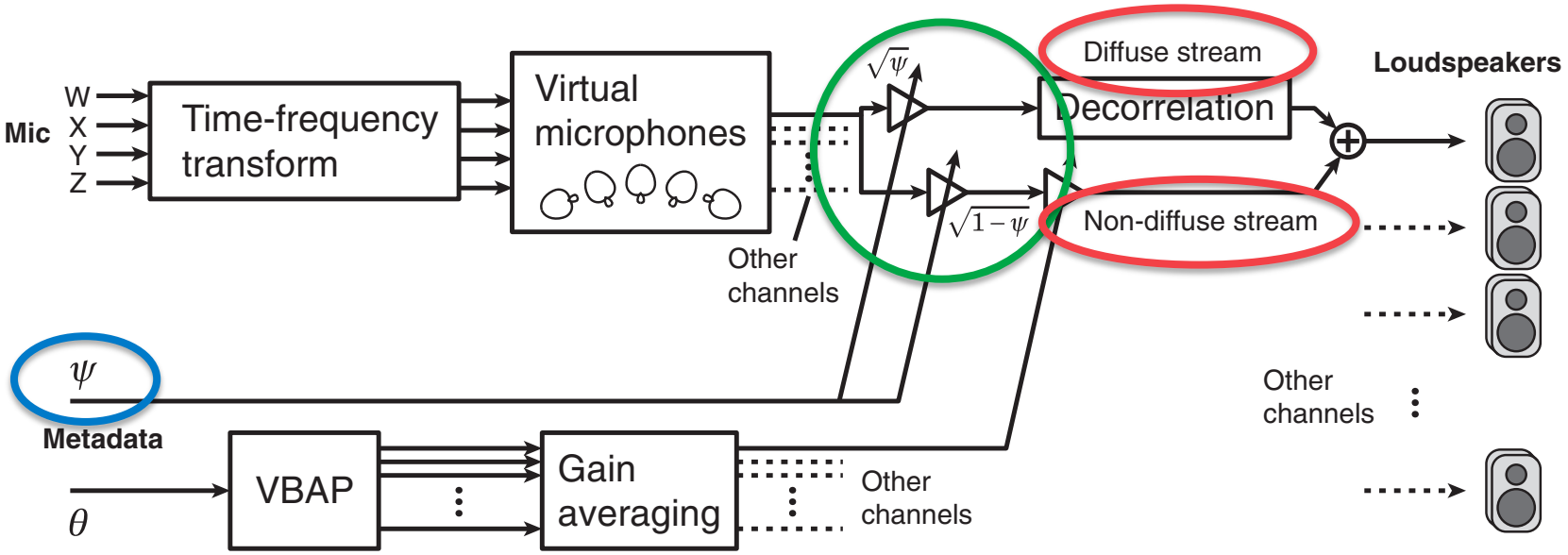


Directional audio coding (DirAC)
Virtual microphones

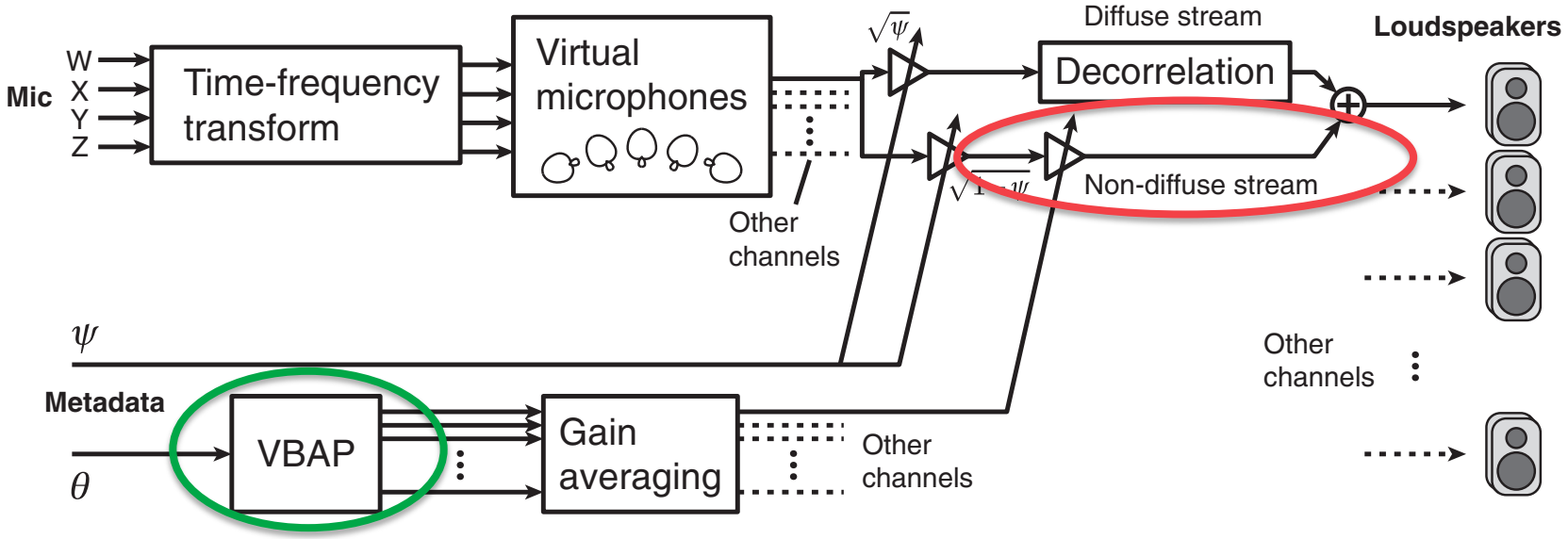
- Virtual microphones are created as a weighted sum of the B-format signals (e.g. supercardioids)
- One microphone points to each loudspeaker:



Directional audio coding (DirAC)
DirAC synthesis



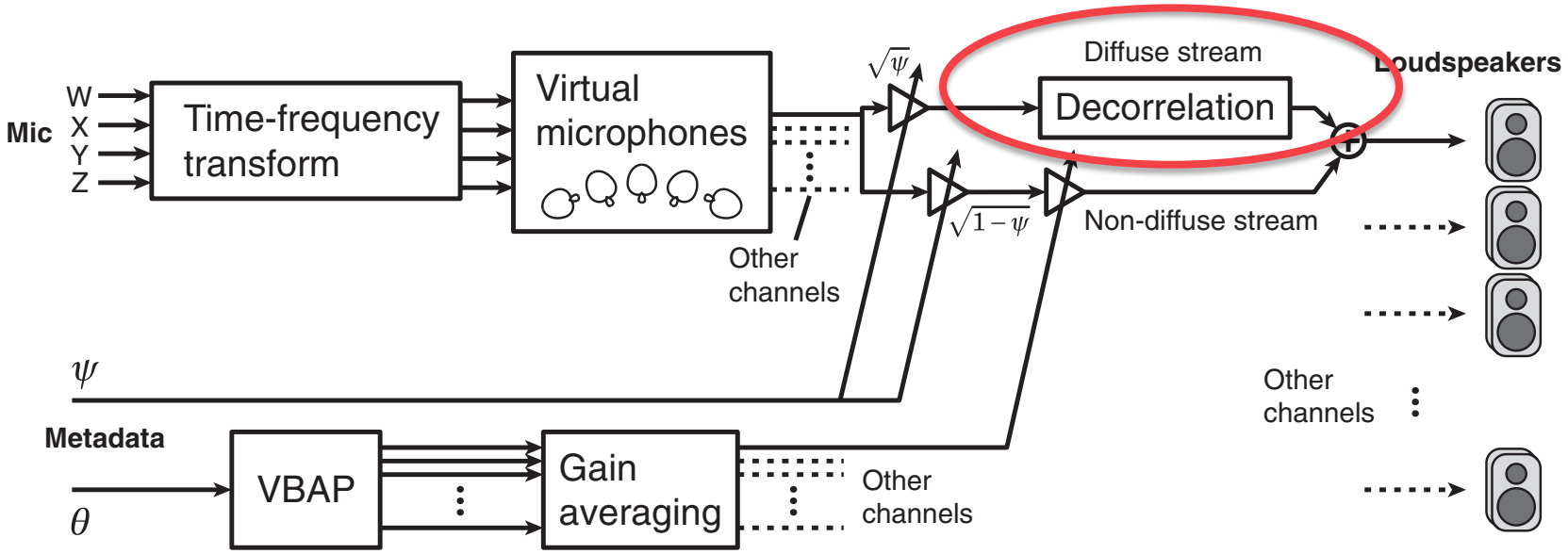
Directional audio coding (DirAC)
DirAC synthesis



Directional audio coding (DirAC)
Nondiffuse stream

- Includes mostly the part of the sound that has a certain direction
  - Reproduced as point-sources
- Vector base amplitude panning (VBAP)
  - According to the direction sent in the metadata
- In many cases the direction in metadata is subject to abrupt temporal changes
  - Gain smoothing is applied to avoid artifacts

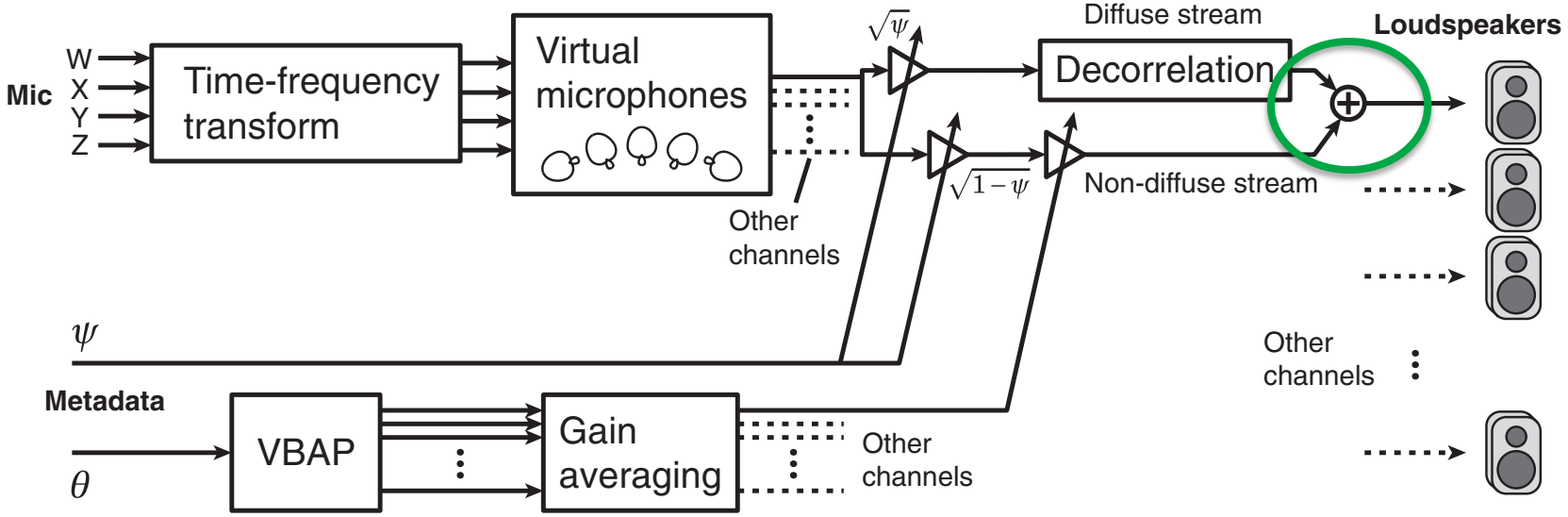
Directional audio coding (DirAC)
DirAC synthesis



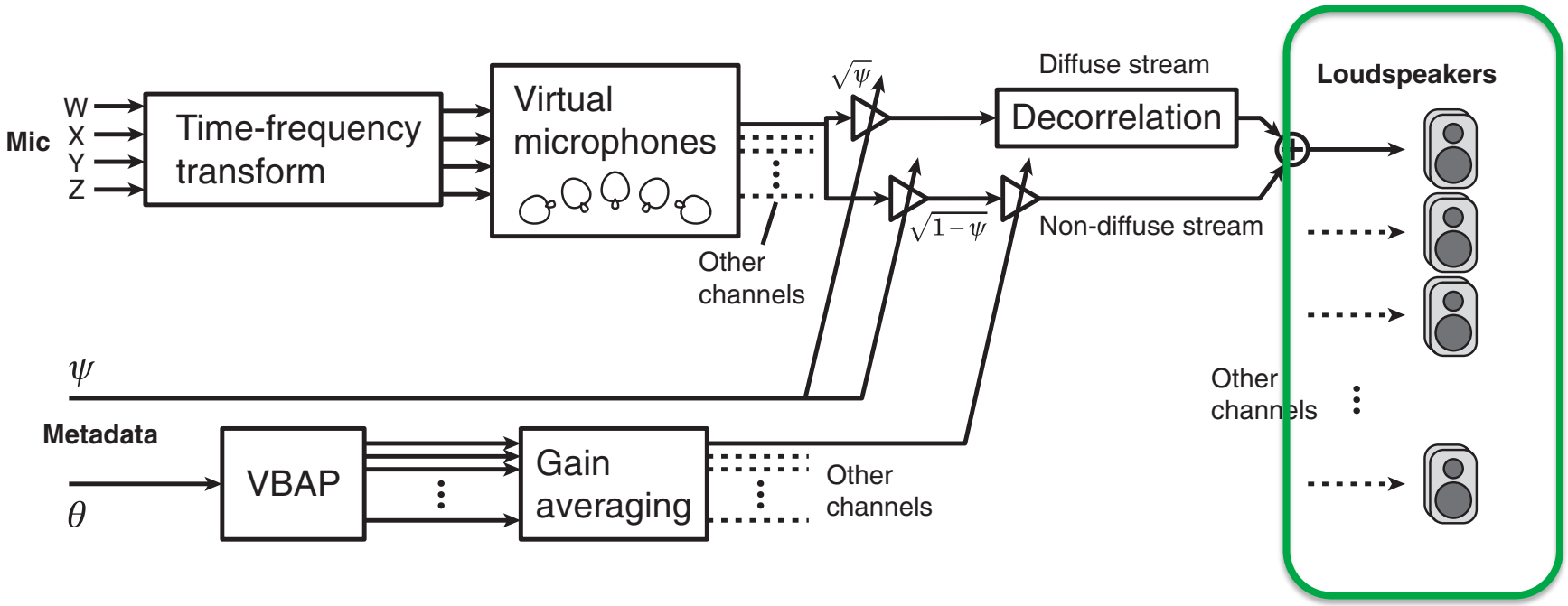
Directional audio coding (DirAC)
Diffuse stream

- Includes mostly the reverberant and the ambient parts
  - Reproduced as surrounding
- Decorrelation
  - E.g. frequency-dependent delays or noise-bursts

Directional audio coding (DirAC)
DirAC synthesis



Directional audio coding (DirAC)
DirAC synthesis



Directional audio coding (DirAC)
Basic idea

- Synthesize features that translate into parameters for spatial-sound perception in hearing:

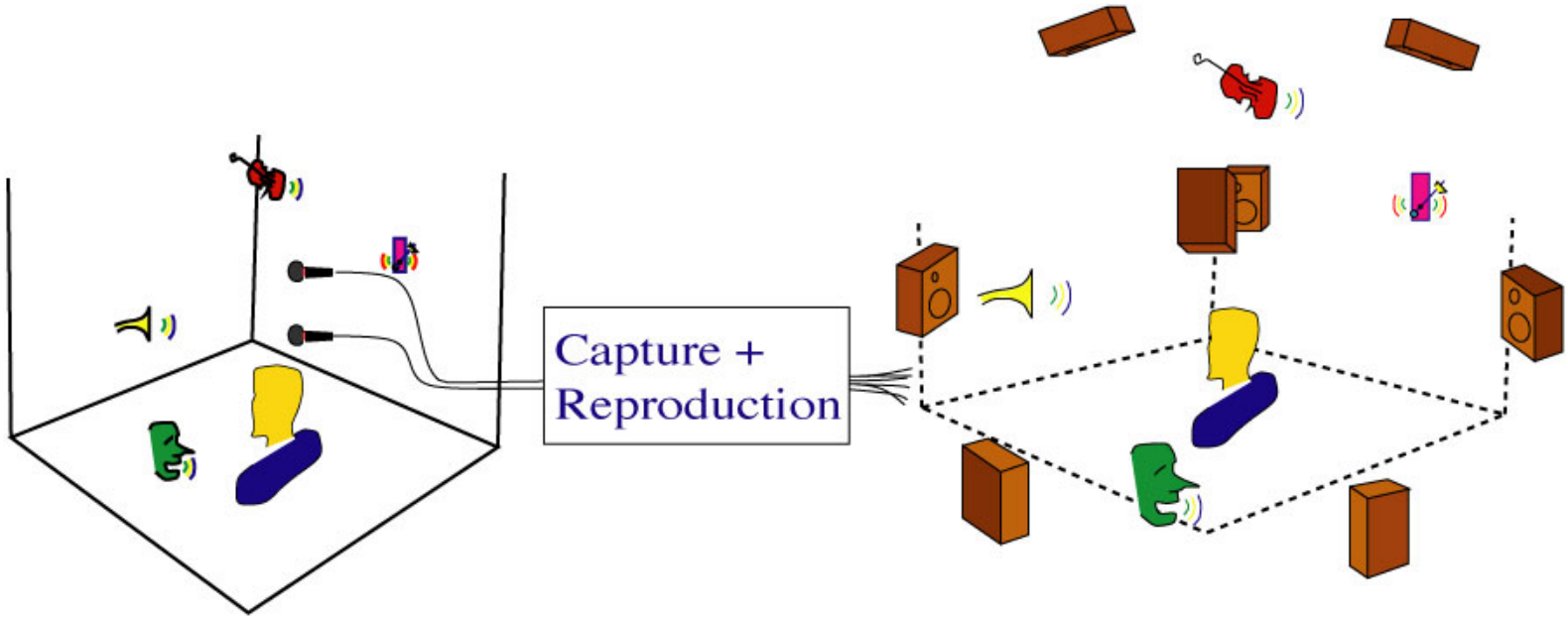
Reproduction		Human hearing
VBAP	→	ITD, ILD
Decorrelation	→	Coherence

Applications

- Spatial sound capturing and reproduction
- Teleconferencing
- Spatial impulse response rendering (SIRR)
- Virtual world / game audio

- Combinations of all these techniques!

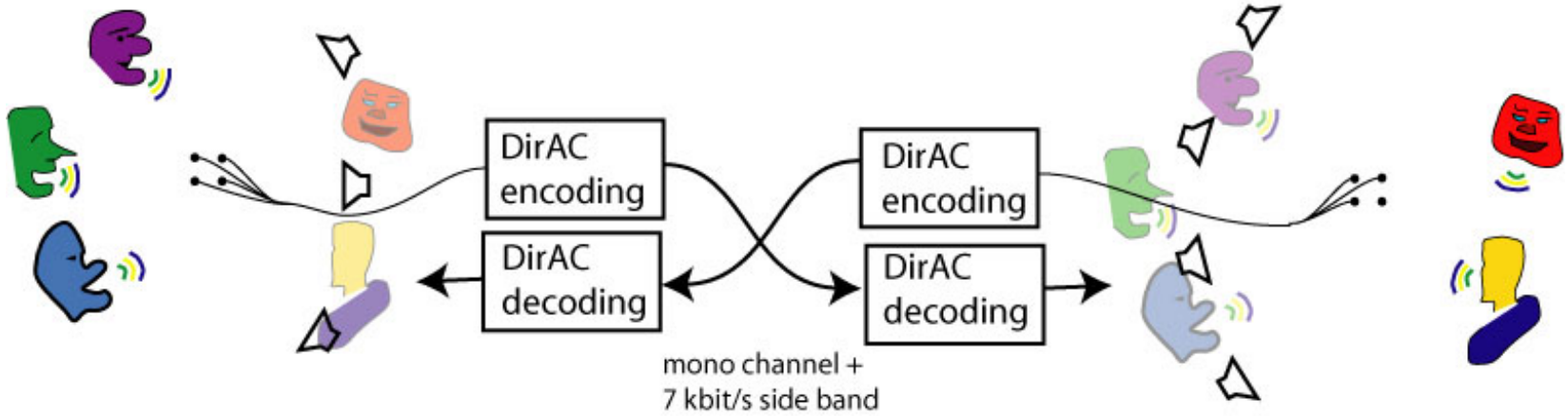
Applications
Spatial sound reproduction



Applications
Spatial sound reproduction

- Spatial sound is captured with a B-format microphone
- DirAC processing can be applied for arbitrary loudspeaker layouts
- Reproduction is also possible with headphones (with or without head tracking)

Applications
Teleconferencing

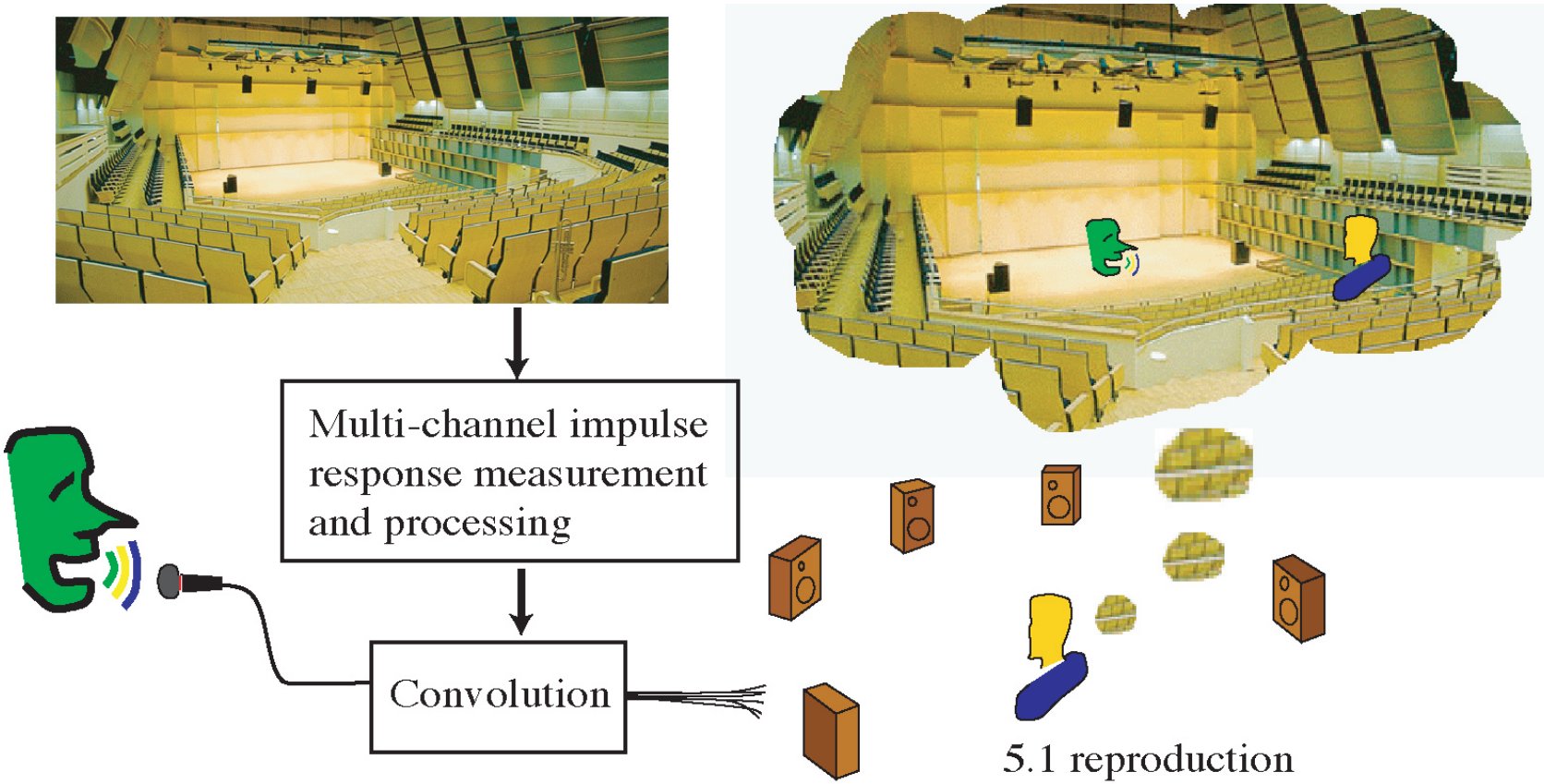


Applications
Teleconferencing

- Only mono channel + low-bit-rate metadata is sent
  - Good quality can still be obtained
- Locations of the participants are reproduced
  - Increases immersion and naturalness
  - Increases speech intelligibility

Applications

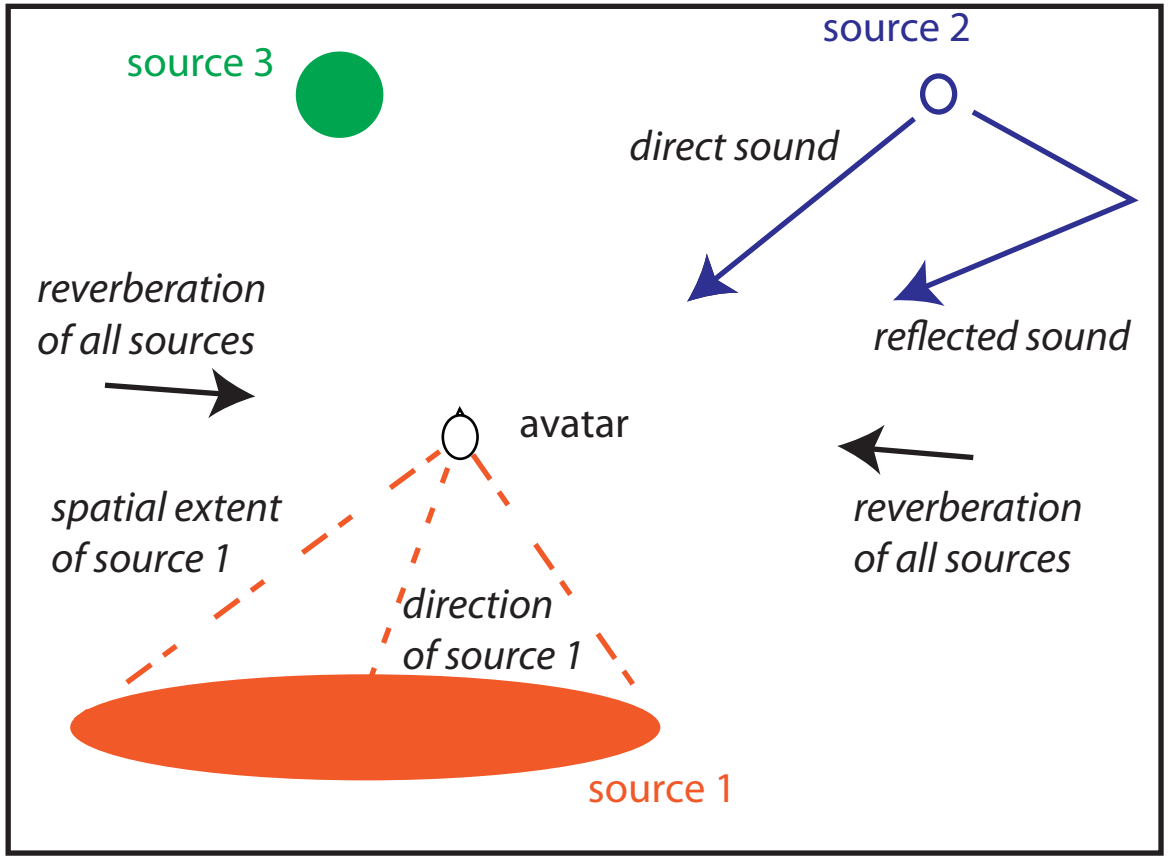
Spatial impulse response rendering



Applications
Spatial impulse response rendering

- Basically the same processing as in DirAC
- Smoothing is turned off
- High temporal resolution is important

Applications
Virtual worlds



Applications
Virtual worlds

- DirAC can perform all traditional tasks for spatial sound reproduction in virtual worlds
  - Source positioning
  - Reverberation generation
- In addition
  - Spatial extent can be controlled efficiently
  - Reverberation can be generated efficiently

Applications
Combining different applications

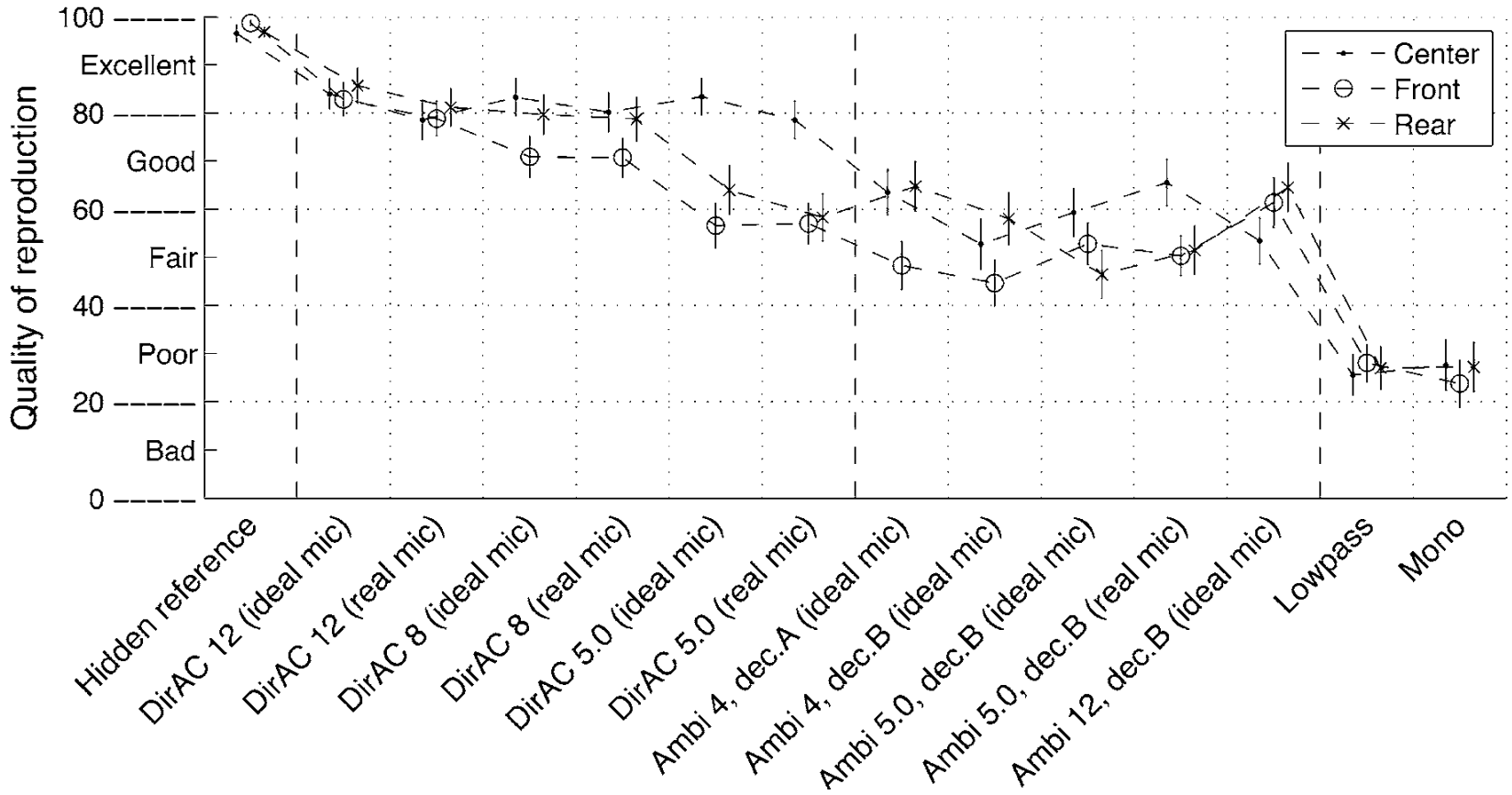
- These methods can also be used at the same time
- Audio from various sources can be processed with the same “backend”
  - No significant increase in the computational complexity

Evaluation

- Signal-dependent nonlinear processing
- Does not target to reconstruct wave field
- Must be evaluated with human listeners

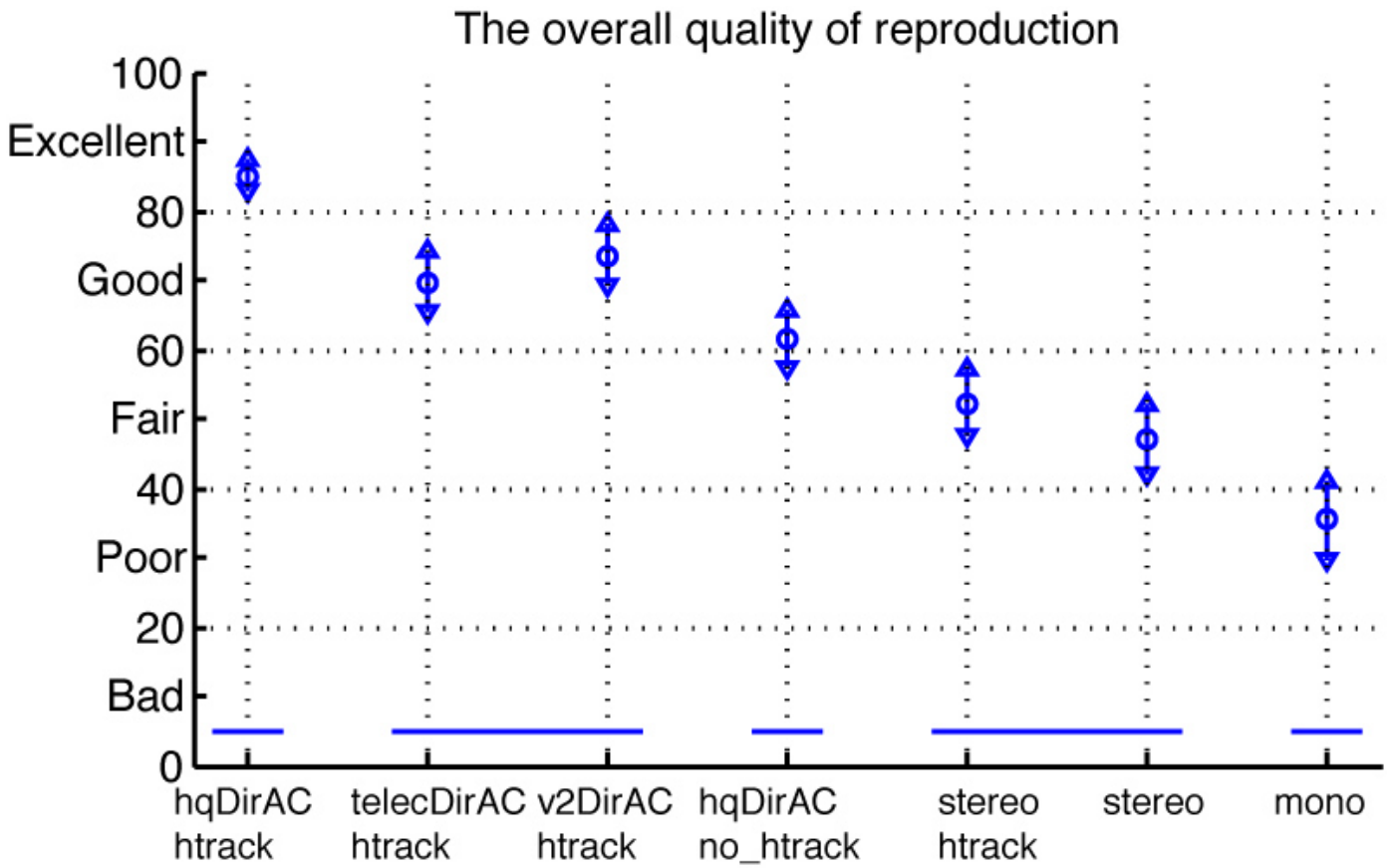
Evaluation

Loudspeaker reproduction

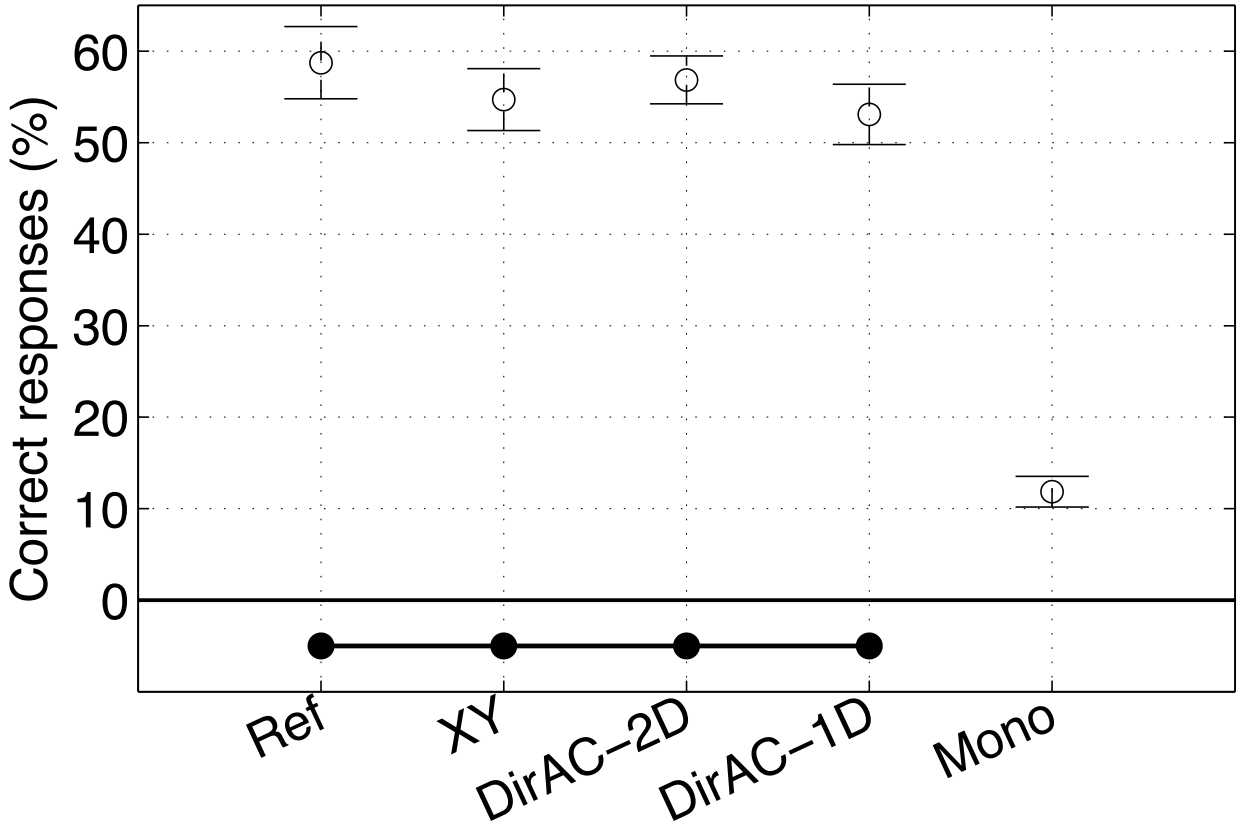


Evaluation

Headphone reproduction

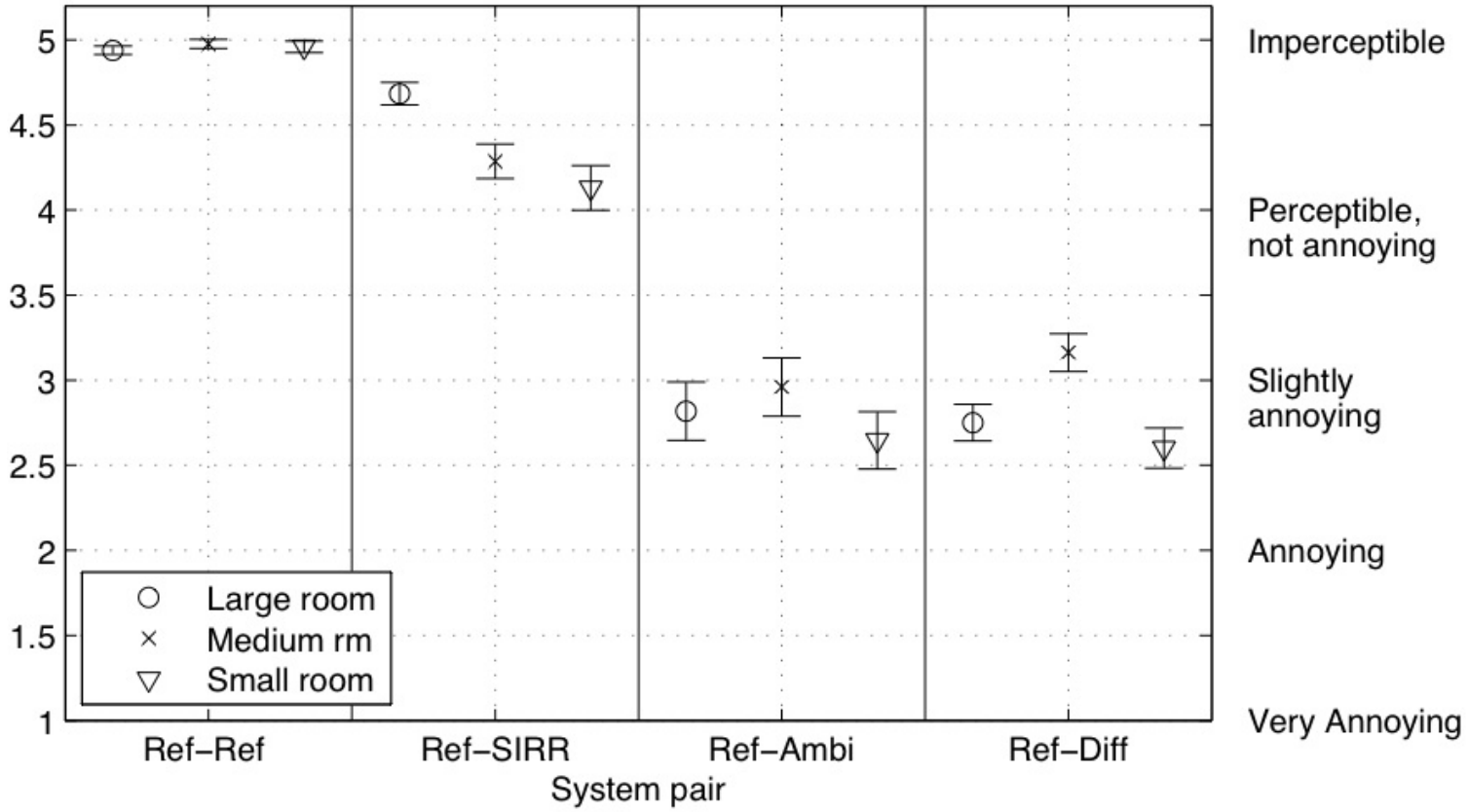


Evaluation
Teleconference



Evaluation

Spatial impulse response rendering



Evaluation Challenges

- Temporal resolution is not infinite
  - Smoothing
- Decorrelation smears the signal in time

Summary

- Perception-based reproduction of spatial audio
- Input from first-order microphones
- Directional analysis of sound field in auditory frequency bands
- Nondiffuse sound reproduced as point sources
- Diffuse sound reproduced as surrounding
- Many applications
- Listening tests: difference to reference condition small

References

- V. Pulkki. **Spatial sound reproduction with directional audio coding**. J. Audio Eng. Soc., 55(6):503–516, June 2007.
- J. Vilkamo, T. Lokki, and V. Pulkki. **Directional audio coding: Virtual microphone based synthesis and subjective evaluation**. J. Audio Eng. Soc., 57(9):709:724, September 2009.
- M.-V. Laitinen and V. Pulkki. **Binaural reproduction for directional audio coding**. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY, U.S.A., 2009.
- J. Ahonen and V. Pulkki. **Speech Intelligibility in Teleconference Application of Directional Audio Coding**. AES 40th international conference, Tokyo, Japan, October 2010.
- V. Pulkki and J. Merimaa. **Spatial impulse response rendering 2: Reproduction of Diffuse Sound and Listening Tests**. J. Audio Eng. Soc., 54(1-2):3–20, February 2006.
- M.-V. Laitinen, T. Pihlajamäki, C. Erkut, and V. Pulkki. **Parametric time-frequency representation of spatial sound in virtual worlds**. ACM Transactions on Applied Perception, June 2012.

Thank you!