FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Current Challenges in Multichannel Acoustic Signal Processing for Natural Human/Machine Interfaces

Walter Kellermann

with contributions from

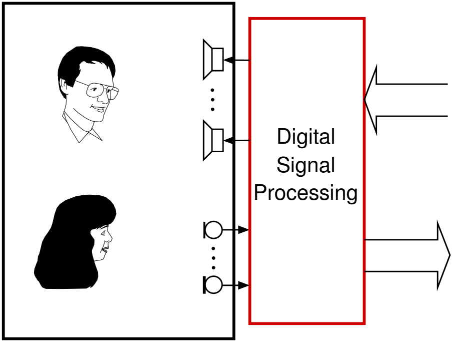
H. Barfuss, A. Brendel, H. Buchner, M. Bürger, A. Deleforge, K. Reindl, A. Schmidt, M. Schneider, A. Schwarz, Y. Zheng, a.o.

Multimedia Communications and Signal Processing

University Erlangen-Nuremberg

Generic Scenario:
Natural Interactive and Immersive Human/Machine Interface

Mobile users, distant microphones/loudspeakers



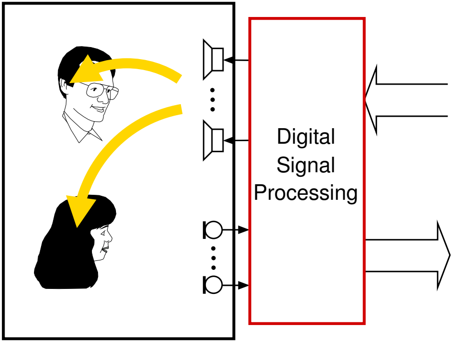
- Rendering - Reproduce desired signals at distant ears
- Acquisition - Localize sources and capture clean signals from distance

- Feedback of loudspeaker signals
- Noise and interferers
- Reverberation

Generic Scenario:

Natural Interactive and Immersive Human/Machine Interface

Mobile users, distant microphones/loudspeakers

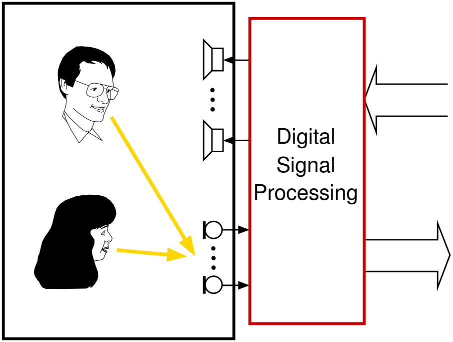


Tasks:

- Rendering - Reproduce desired signals at distant ears
- Acquisition - Localize sources and capture clean signals from distance
- Feedback of loudspeaker signals
- Noise and interferers
- Reverberation

Generic Scenario:
Natural Interactive and Immersive Human/Machine Interface

Mobile users, distant microphones/loudspeakers

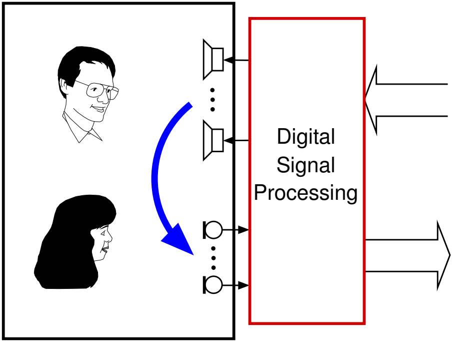


Tasks:

- Rendering - Reproduce desired signals at distant ears
- Acquisition - Localize sources and capture clean signals from distance
- Feedback of loudspeaker signals
- Noise and interferers
- Reverberation

Generic Scenario:
Natural Interactive and Immersive Human/Machine Interface

Mobile users, distant microphones/loudspeakers



Tasks:

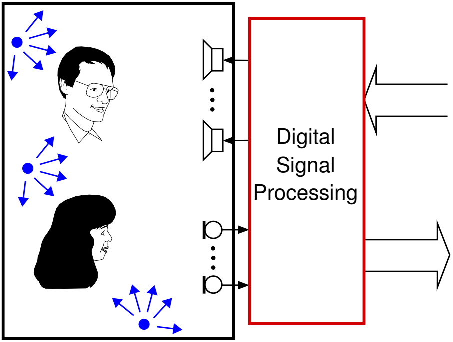
- Rendering - Reproduce desired signals at distant ears
- Acquisition - Localize sources and capture clean signals from distance

Challenges:

- Feedback of loudspeaker signals
- Noise and interferers
- Reverberation

Generic Scenario:
Natural Interactive and Immersive Human/Machine Interface

Mobile users, distant microphones/loudspeakers



Tasks:

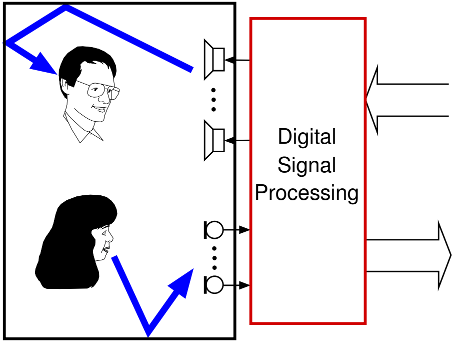
- Rendering - Reproduce desired signals at distant ears
- Acquisition - Localize sources and capture clean signals from distance

Challenges:

- Feedback of loudspeaker signals
- Noise and interferers
- Reverberation

Generic Scenario:
Natural Interactive and Immersive Human/Machine Interface

Mobile users, distant microphones/loudspeakers



Tasks:

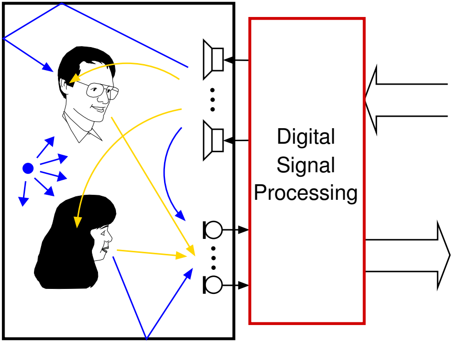
- Rendering - Reproduce desired signals at distant ears
- Acquisition - Localize sources and capture clean signals from distance

Challenges:

- Feedback of loudspeaker signals
- Noise and interferers
- Reverberation

Generic Scenario:
Natural Interactive and Immersive Human/Machine Interface

Mobile users, distant microphones/loudspeakers



Tasks:

- Rendering - Reproduce desired signals at distant ears
- Acquisition - Localize sources and capture clean signals from distance

Challenges:

- Feedback of loudspeaker signals
- Noise and interferers
- Reverberation

Applications

- **'Far-field' Communication and Human/Machine Interfaces** for
  - mobile phones ('smartphones'), mobile computing devices (tablets), smart watches
  - desktop computers, info-/edutainment terminals, game stations
  - telepresence systems (offices,. . ., classrooms, . . ., auditoria)
  - car interiors (with, e.g., in-car communication, seat-specific rendering . . .)
  - ambient communication (smart meeting rooms, smart homes, museums and exhibitions, . . .)
- **Audio Communication**
  - equipment for stages and recording studios
  - virtual acoustic environments (virtual concert halls, teleteaching studios,. . .)
- **Safety and Surveillance**
  - acoustic displays in control centers, cockpits
  - acoustic monitoring in health care, elderly care
  - acoustic scene analysis (train stations, . . .)

Applications

- **'Far-field' Communication and Human/Machine Interfaces** for
  - mobile phones ('smartphones'), mobile computing devices (tablets), smart watches
  - desktop computers, info-/edutainment terminals, game stations
  - telepresence systems (offices,. . ., classrooms, . . ., auditoria)
  - car interiors (with, e.g., in-car communication, seat-specific rendering . . .)
  - ambient communication (smart meeting rooms, smart homes, museums and exhibitions, . . .)
- **Audio Communication**
  - equipment for stages and recording studios
  - virtual acoustic environments (virtual concert halls, teleteaching studios,. . .)
- **Safety and Surveillance**
  - acoustic displays in control centers, cockpits
  - acoustic monitoring in health care, elderly care
  - acoustic scene analysis (train stations, . . .)

Applications

- **'Far-field' Communication and Human/Machine Interfaces** for
  - mobile phones ('smartphones'), mobile computing devices (tablets), smart watches
  - desktop computers, info-/edutainment terminals, game stations
  - telepresence systems (offices,. . ., classrooms, . . ., auditoria)
  - car interiors (with, e.g., in-car communication, seat-specific rendering . . .)
  - ambient communication (smart meeting rooms, smart homes, museums and exhibitions, . . .)
- **Audio Communication**
  - equipment for stages and recording studios
  - virtual acoustic environments (virtual concert halls, teleteaching studios,. . .)
- **Safety and Surveillance**
  - acoustic displays in control centers, cockpits
  - acoustic monitoring in health care, elderly care
  - acoustic scene analysis (train stations, . . .)

Applications

- **'Far-field' Communication and Human/Machine Interfaces** for
  - mobile phones ('smartphones'), mobile computing devices (tablets), smart watches
  - desktop computers, info-/edutainment terminals, game stations
  - telepresence systems (offices,. . ., classrooms, . . ., auditoria)
  - car interiors (with, e.g., in-car communication, seat-specific rendering . . .)
  - ambient communication (smart meeting rooms, smart homes, museums and exhibitions, . . .)
- **Audio Communication**
  - equipment for stages and recording studios
  - virtual acoustic environments (virtual concert halls, teleteaching studios,. . .)
- **Safety and Surveillance**
  - acoustic displays in control centers, cockpits
  - acoustic monitoring in health care, elderly care
  - acoustic scene analysis (train stations, . . .)

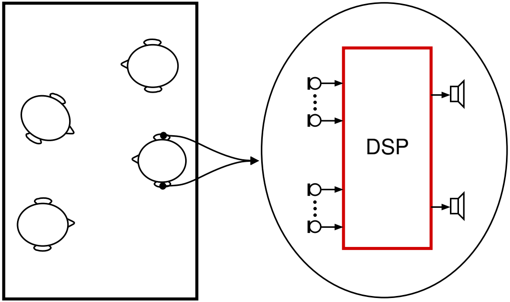
Applications (cont'd)

- **Autonomous systems**, e.g., Robot Audition

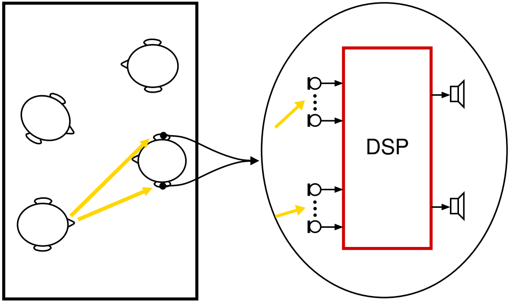
Source: ieee.spectrum.org, 2012

- **Acoustic sensor networks** supporting the applications above ...

Another Immersive Scenario: 'Listening devices'



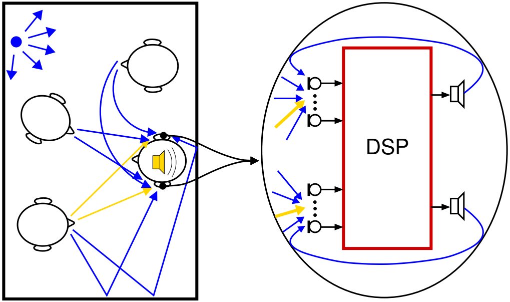
Another Immersive Scenario: 'Listening devices'



Tasks:

- Rendering - Reproduce undistorted signals with binaural cues
- Acquisition - Localize desired source(s) and enhance desired signal(s)

Another Immersive Scenario: 'Listening devices'



Tasks:

- Rendering - Reproduce undistorted signals with binaural cues
- Acquisition - Localize desired source(s) and enhance desired signal(s)

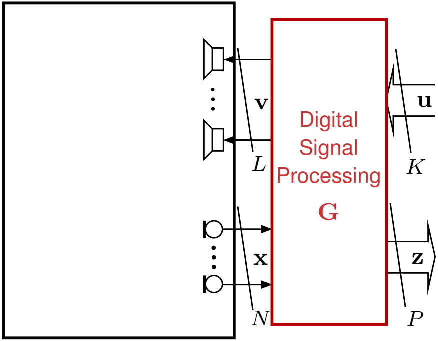
Challenges:

- Loudspeaker feedback (howling)
- Noise and interferers
- Reverberation

Applications

- **Hearing aids**
- **Headphones**, e.g., for
  - hearing protection in noisy environments (construction work, mining,. . .)
  - active noise cancellation systems
  - mobile phones, mobile computing devices, etc.
  - augmented/virtual reality
  - ’Smart Headphones’/’Hearables’,
- . . . combined with **Acoustic Sensor Networks**

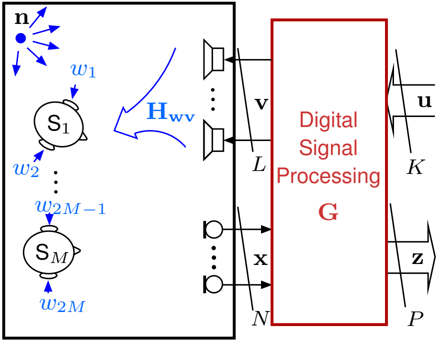
Signal Model



Linear MIMO system $\mathbf{G}$:

$$\begin{pmatrix} \mathbf{v} \\ \mathbf{z} \end{pmatrix} = \mathbf{G} * \begin{pmatrix} \mathbf{u} \\ \mathbf{x} \end{pmatrix} = \begin{pmatrix} \mathbf{G_{vu}} & \mathbf{G_{vx}} \\ \mathbf{G_{zu}} & \mathbf{G_{zx}} \end{pmatrix} * \begin{pmatrix} \mathbf{u} \\ \mathbf{x} \end{pmatrix}$$

Signal Model



Linear MIMO system $\mathbf{G}$:

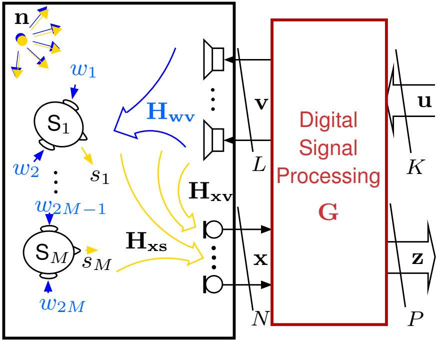
$$\begin{pmatrix} \mathbf{v} \\ \mathbf{z} \end{pmatrix} = \mathbf{G} * \begin{pmatrix} \mathbf{u} \\ \mathbf{x} \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{G_{vu}} & \mathbf{G_{vx}} \\ \mathbf{G_{zu}} & \mathbf{G_{zx}} \end{pmatrix} * \begin{pmatrix} \mathbf{u} \\ \mathbf{x} \end{pmatrix}$$

Listeners' signals:

$$\mathbf{w} = \mathbf{H_{wv}} * \mathbf{v} + \mathbf{n_w}$$

Signal Model



Linear MIMO system $\mathbf{G}$:

$$\begin{pmatrix} \mathbf{v} \\ \mathbf{z} \end{pmatrix} = \mathbf{G} * \begin{pmatrix} \mathbf{u} \\ \mathbf{x} \end{pmatrix} = \begin{pmatrix} \mathbf{G_{vu}} & \mathbf{G_{vx}} \\ \mathbf{G_{zu}} & \mathbf{G_{zx}} \end{pmatrix} * \begin{pmatrix} \mathbf{u} \\ \mathbf{x} \end{pmatrix}$$

Listeners' signals:

$$\mathbf{w} = \mathbf{H_{wv}} * \mathbf{v} + \mathbf{n_w}$$

Microphone signals:

$$\mathbf{x} = \mathbf{H_{xs}} * \mathbf{s} + \mathbf{H_{xv}} * \mathbf{v} + \mathbf{n_x}$$

Signal Model (cont'd)

Elements of $\mathbf{H}_{\mathbf{wv}}, \mathbf{H}_{\mathbf{xv}}, \mathbf{H}_{\mathbf{xs}}$ are impulse responses.

- **Reverberation time T_{60}**
  (sound energy decayed by 60dB)
  - car ≈ 50ms
  - concert halls $\approx 1 \dots 2$s
- **FIR models**
  - typically $L_H \approx T_{60} \cdot f_s/3$ coefficients
  - nonminimum-phase
  - many zeros close to unit circle

Signal Model (cont'd)

Elements of $\mathbf{H}_{\mathbf{wv}}, \mathbf{H}_{\mathbf{xv}}, \mathbf{H}_{\mathbf{xs}}$ are impulse responses.

- **Reverberation time T_{60}**
  (sound energy decayed by 60dB)
  - car ≈ 50ms
  - concert halls $\approx 1 \ldots 2$s
- **FIR models**
  - typically $L_H \approx T_{60} \cdot f_s/3$ coefficients
  - nonminimum-phase
  - many zeros close to unit circle

Signal Model (cont'd)

Elements of $\mathbf{H}_{\mathbf{wv}}, \mathbf{H}_{\mathbf{xv}}, \mathbf{H}_{\mathbf{xs}}$ are impulse responses.

- **Reverberation time T_{60}**

  (sound energy decayed by 60dB)
  - ▷ car ≈ 50ms
  - ▷ concert halls $\approx 1 \ldots 2$s

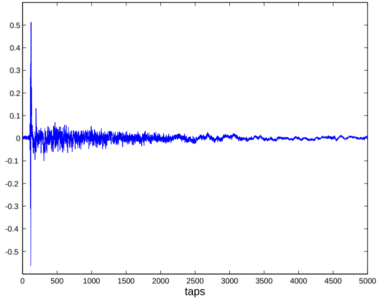
- **FIR models**
  - ▷ typically $L_H \approx T_{60} \cdot f_s/3$ coefficients
  - ▷ nonminimum-phase
  - ▷ many zeros close to unit circle

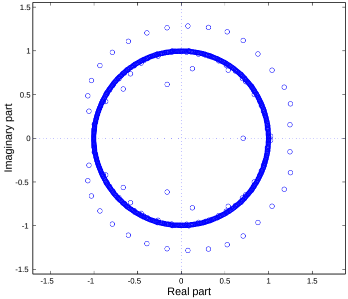
Signal Model (cont'd)

Elements of $\mathbf{H}_{\mathbf{wv}}, \mathbf{H}_{\mathbf{xv}}, \mathbf{H}_{\mathbf{xs}}$ are impulse responses.

- **Reverberation time T_{60}**
  (sound energy decayed by 60dB)
  - car ≈ 50ms
  - concert halls $\approx 1 \ldots 2$s
- **FIR models**
  - typically $L_H \approx T_{60} \cdot f_s / 3$ coefficients
  - nonminimum-phase
  - many zeros close to unit circle

Example: Office 5.5m $\times$ 3m $\times$ 2.8m, $T_{60} \approx 300$msec, sampling $f_s = 12$kHz.





Overview

Reproduction

- Binaural rendering
- Multi-zone sound field synthesis

Acquisition

Summary and Outlook

Overview

Reproduction

Acquisition

- MIMO acoustic echo cancellation
- Informed signal extraction
- Blind signal extraction
- Localization
- Interference and noise estimation
- Robot audition

Summary and Outlook

Overview

Reproduction

Acquisition

Summary and Outlook

Overview

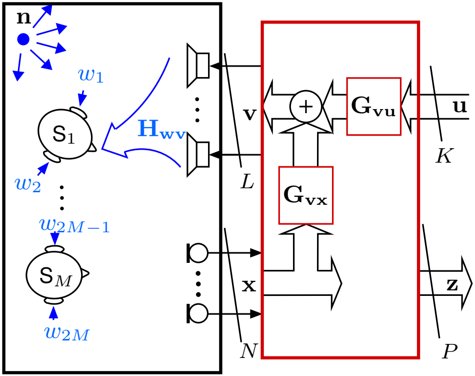
Reproduction

- Binaural rendering
- Multi-zone sound field synthesis

Acquisition

Summary and Outlook

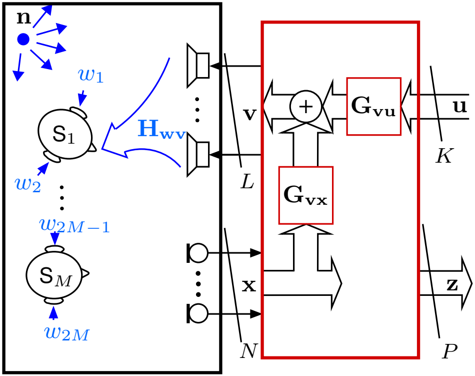
Fundamental Problems for Reproduction



Desired signals $\mathrm{w_d}$:

$$\mathbf{w}_\mathrm{d} \stackrel{!}{=} \mathbf{H}_\mathrm{d} * \mathbf{u}$$

Fundamental Problems for Reproduction



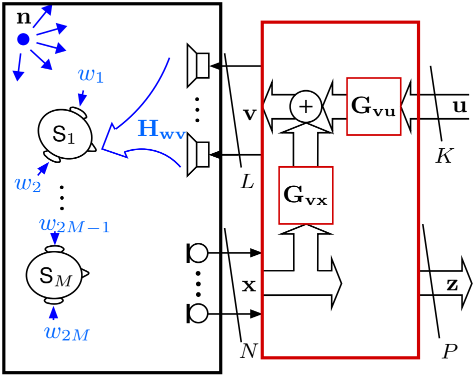
Desired signals $\mathrm{w_d}$:

$$\mathbf{w}_\mathrm{d} \overset{!}{=} \mathbf{H}_\mathrm{d} * \mathbf{u}$$

$\Rightarrow$ Requirement:

$$\mathbf{H}_\mathrm{d} * \mathbf{u} \overset{!}{=} \mathbf{H}_\mathbf{wv} * (\mathbf{G}_\mathbf{vu} * \mathbf{u} + \mathbf{G}_\mathbf{vx} * \mathbf{x}) + \mathbf{n}_\mathbf{w}$$

Fundamental Problems for Reproduction



Desired signals $\mathbf{w}_\mathrm{d}$:

$$\mathbf{w}_\mathrm{d} \stackrel{!}{=} \mathbf{H}_\mathrm{d} * \mathbf{u}$$

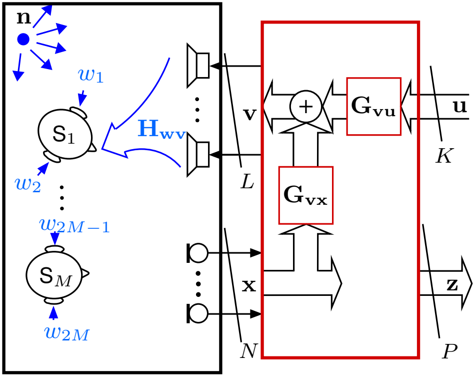
$\Rightarrow$ Requirement:

$$\mathbf{H}_\mathrm{d} * \mathbf{u} \stackrel{!}{=} \mathbf{H}_{\mathbf{wv}} * (\mathbf{G}_{\mathbf{vu}} * \mathbf{u} + \mathbf{G}_{\mathbf{vx}} * \mathbf{x}) + \mathbf{n}_\mathbf{w}$$

$\Rightarrow$ **2 Subproblems:**

- Equalization (Deconvolution):
  $$\mathbf{H}_{\mathbf{wv}} * \mathbf{G}_{\mathbf{vu}} * \mathbf{u} \stackrel{!}{=} \mathbf{H}_\mathrm{d} * \mathbf{u}.$$
- Interference compensation:
  $$\mathbf{H}_{\mathbf{wv}} * \mathbf{G}_{\mathbf{vx}} * \mathbf{x} + \mathbf{n}_\mathbf{w} \stackrel{!}{=} \mathbf{0}$$

Fundamental Problems for Reproduction



Desired signals $\mathrm{w_d}$:

$$\mathbf{w}_\mathrm{d} \stackrel{!}{=} \mathbf{H}_\mathrm{d} * \mathbf{u}$$

$\Rightarrow$ Requirement:

$$\mathbf{H}_\mathrm{d} * \mathbf{u} \stackrel{!}{=} \mathbf{H}_{\mathbf{wv}} * (\mathbf{G}_{\mathbf{vu}} * \mathbf{u} + \mathbf{G}_{\mathbf{vx}} * \mathbf{x}) + \mathbf{n}_\mathbf{w}$$

$\Rightarrow$ **2 Subproblems:**

- Equalization (Deconvolution):

$$\mathbf{H}_{\mathbf{wv}} * \mathbf{G}_{\mathbf{vu}} * \mathbf{u} \stackrel{!}{=} \mathbf{H}_\mathrm{d} * \mathbf{u}.$$

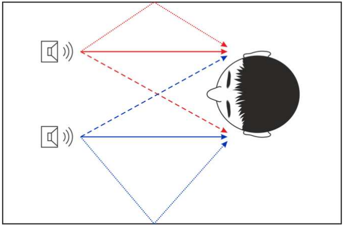
- Interference compensation:

$$\mathbf{H}_{\mathbf{wv}} * \mathbf{G}_{\mathbf{vx}} * \mathbf{x} + \mathbf{n}_\mathbf{w} \stackrel{!}{=} \mathbf{0}$$

Challenges: Blind equalization of $\mathbf{H}_{\mathbf{wv}}$ and **estimation (prediction)** of $\mathbf{n}_\mathbf{w}$!

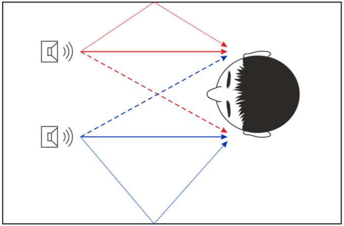
State of the Art in Reproduction I

- **Binaural Rendering** (*Hamada et al. 1998*; *Poletti 2008*)



State of the Art in Reproduction I

- **Binaural Rendering** (*Hamada et al. 1998*; *Poletti 2008*)



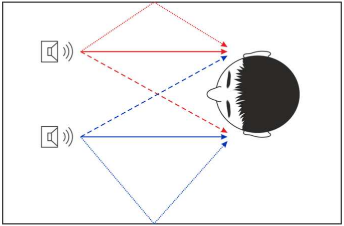
▷ **Ideally:** perfect cross-talk cancellation (CTC) and room equalization

$$\mathbf{H_{wv}} * \mathbf{G_{vu}} = \mathbf{I}$$

with known personalized Head-Related Impulse Responses HRIRs $\mathbf{H_{wv}}$ in time-invariant listening scenarios

State of the Art in Reproduction I

- **Binaural Rendering** (*Hamada et al. 1998*; *Poletti 2008*)



- ▷ **Ideally:** perfect cross-talk cancellation (CTC) and room equalization

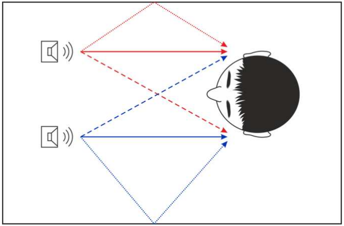
$$\mathbf{H_{wv}} * \mathbf{G_{vu}} = \mathbf{I}$$

  with known personalized Head-Related Impulse Responses HRIRs $\mathbf{H_{wv}}$ in time-invariant listening scenarios

- ▷ **In practice:** nonrobust to head movements, sensitive to errors in HRIRs, ill-conditioning increases for multiple listeners

State of the Art in Reproduction I

- **Binaural Rendering** (*Hamada et al. 1998*; *Poletti 2008*)



- ▷ **Ideally:** perfect cross-talk cancellation (CTC) and room equalization

$$\mathbf{H_{wv}} * \mathbf{G_{vu}} = \mathbf{I}$$

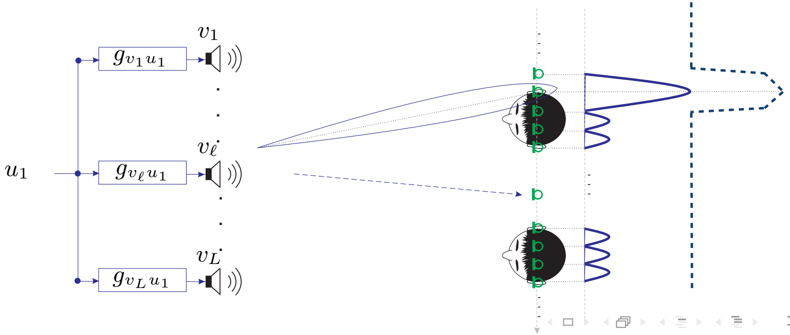
with known personalized Head-Related Impulse Responses HRIRs $\mathbf{H_{wv}}$ in time-invariant listening scenarios

- ▷ **In practice:** nonrobust to head movements, sensitive to errors in HRIRs, ill-conditioning increases for multiple listeners
- ▷ No suppression of local noise and interference: $\mathbf{G_{vx}} = \mathbf{0}$

Recent Developments in Reproduction I

- **Personalized Binaural Beamforming**
  - Crosstalk suppression for binauralization and personalization
  - Coarse room equalization for sweet spots around ears (if known or modelled) with impulse responses as elements of $\mathbf{H}_\mathrm{d}$ to obtain

$$\mathbf{H}_\mathbf{wv} * \mathbf{G}_\mathbf{vu} \approx \mathbf{H}_\mathrm{d}$$

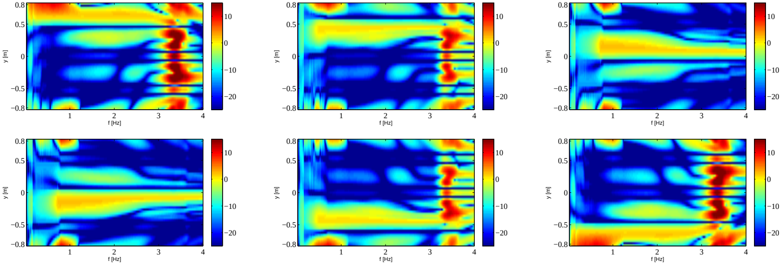


Recent Developments in Reproduction I

Personalized Binaural Beamforming - Example:

$L = 8$ loudspeakers (spacing 15cm)

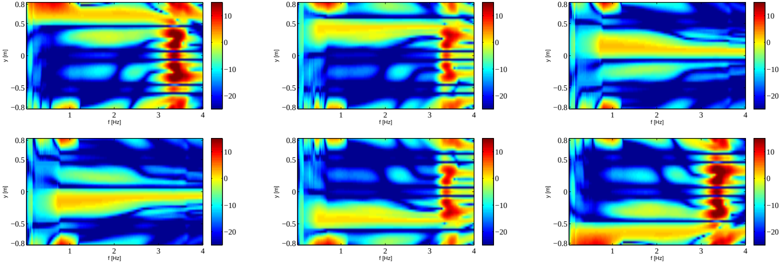
$M = 6$ ears (3 listeners, spacing 50cm, at 1.5m distance)



Recent Developments in Reproduction I

Personalized Binaural Beamforming - Example:

$L = 8$ loudspeakers (spacing 15cm)
$M = 6$ ears (3 listeners, spacing 50cm, at 1.5m distance)



- More robust than CTC to head movements and variability of HRIRs
- No suppression of local noise and interference: $\mathbf{G_{vx}} = \mathbf{0}$

State of the Art in Reproduction II

- **Sound field synthesis** ($\rightarrow$ Wavefield Synthesis (*TU Delft, 1993ff*), Ambisonics)
  - Concepts for interference compensation (*Kuntz/Rabenstein 2004ff, ..., Sun 2011*) so far limited to time-invariant room acoustics, otherwise $\mathbf{G_{vx}} = \mathbf{0}$
  - Equalization of time-invariant room acoustics (*Spors, 2004ff*; *Gonzalez, 2005ff*; *Talagala, 2015*; *Hofmann 2016*)
  - Adaptive room equalization (*Schneider, 2012*; *Hofmann, 2016*)

State of the Art in Reproduction II

- **Sound field synthesis** (→ Wavefield Synthesis (*TU Delft, 1993ff*), Ambisonics)
  - Concepts for interference compensation (*Kuntz/Rabenstein 2004ff, ..., Sun 2011*) so far limited to time-invariant room acoustics, otherwise $\mathbf{G_{vx}} = \mathbf{0}$
  - Equalization of time-invariant room acoustics (*Spors, 2004ff*; *Gonzalez, 2005ff*; *Talagala, 2015*; *Hofmann 2016*)
  - Adaptive room equalization (*Schneider, 2012*; *Hofmann, 2016*)

State of the Art in Reproduction II

- **Sound field synthesis** (→ Wavefield Synthesis (*TU Delft, 1993ff*), Ambisonics)
  - ▷ Concepts for interference compensation (*Kuntz/Rabenstein 2004ff, ..., Sun 2011*) so far limited to time-invariant room acoustics, otherwise $\mathbf{G_{vx}} = \mathbf{0}$
  - ▷ Equalization of time-invariant room acoustics (*Spors, 2004ff*; *Gonzalez, 2005ff*; *Talagala, 2015*; *Hofmann 2016*)
  - ▷ Adaptive room equalization (*Schneider, 2012*; *Hofmann, 2016*)

State of the Art in Reproduction II

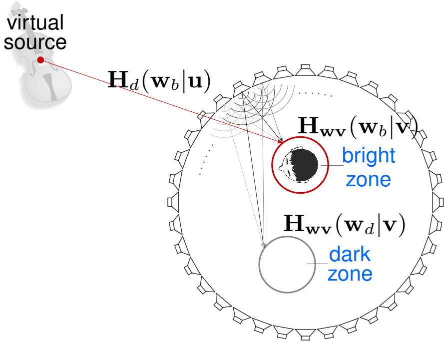
- **Sound field synthesis** (→ Wavefield Synthesis (*TU Delft, 1993ff*), Ambisonics)
  - Concepts for interference compensation (*Kuntz/Rabenstein 2004ff, ..., Sun 2011*) so far limited to time-invariant room acoustics, otherwise $\mathbf{G}_{\mathbf{vx}} = \mathbf{0}$
  - Equalization of time-invariant room acoustics (*Spors, 2004ff*; *Gonzalez, 2005ff*; *Talagala, 2015*; *Hofmann 2016*)
  - Adaptive room equalization (*Schneider, 2012*; *Hofmann, 2016*)

State of the Art in Reproduction II

- **Sound field synthesis** (→ Wavefield Synthesis (*TU Delft, 1993ff*), Ambisonics)
  - Concepts for interference compensation (*Kuntz/Rabenstein 2004ff, ..., Sun 2011*) so far limited to time-invariant room acoustics, otherwise $\mathbf{G_{vx}} = \mathbf{0}$
  - Equalization of time-invariant room acoustics (*Spors, 2004ff*; *Gonzalez, 2005ff*; *Talagala, 2015*; *Hofmann 2016*)
  - Adaptive room equalization (*Schneider, 2012*; *Hofmann, 2016*)

Recent Developments in Reproduction II

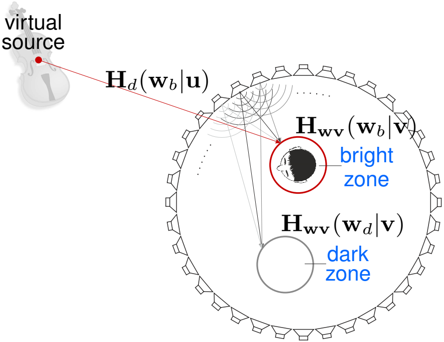
- **Multi-zone Sound Field Synthesis** (*Spors 2008, Abhayapala 2009, Jin 2013, Fazi 2010ff, Buerger 2013ff*)



- ▷ Local sound field synthesis of 'bright zones' and 'dark zones' . . .
- ▷ based on Kirchhoff-Helmholtz integral: requires transfer functions from all loudspeakers to sampling points of enclosing contours and . . .
- ▷ can equalize room acoustics

Recent Developments in Reproduction II

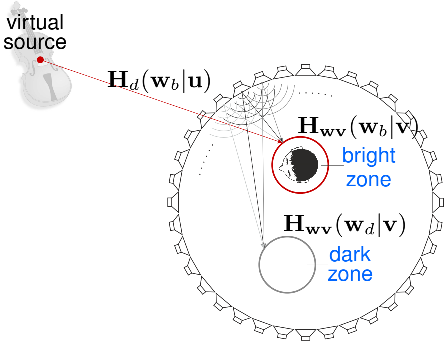
- **Multi-zone Sound Field Synthesis** (*Spors 2008, Abhayapala 2009, Jin 2013, Fazi 2010ff, Buerger 2013ff*)



- Local sound field synthesis of 'bright zones' and 'dark zones' . . .
- based on Kirchhoff-Helmholtz integral: requires transfer functions from all loudspeakers to sampling points of enclosing contours and . . .
- can equalize room acoustics

Recent Developments in Reproduction II

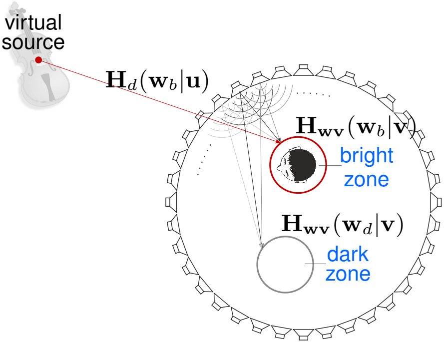
- **Multi-zone Sound Field Synthesis** (*Spors 2008, Abhayapala 2009, Jin 2013, Fazi 2010ff, Buerger 2013ff*)



- ▷ Local sound field synthesis of 'bright zones' and 'dark zones' . . .
- ▷ based on Kirchhoff-Helmholtz integral: requires transfer functions from all loudspeakers to sampling points of enclosing contours and . . .
- ▷ can equalize room acoustics

Recent Developments in Reproduction II

- **Multi-zone Sound Field Synthesis** (*Spors 2008, Abhayapala 2009, Jin 2013, Fazi 2010ff, Buerger 2013ff*)



- ▷ Local sound field synthesis of 'bright zones' and 'dark zones' ...
- ▷ based on Kirchhoff-Helmholtz integral: requires transfer functions from all loudspeakers to sampling points of enclosing contours and ...
- ▷ can equalize room acoustics

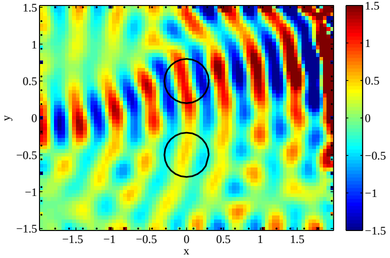
Recent Developments in Reproduction III

Multi-zone Sound Reproduction in Reverberant Environments - Example

(*Buerger 2014*)

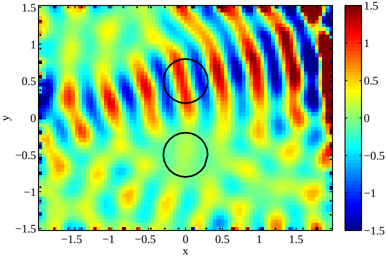
- $L = 70$ loudspeakers in rectangular room;
- image source model (*Allen, 1979*): order 10, reflection coefficient 0.7

No room equalization:



→ Reflections in 'dark zone'

With room equalization:



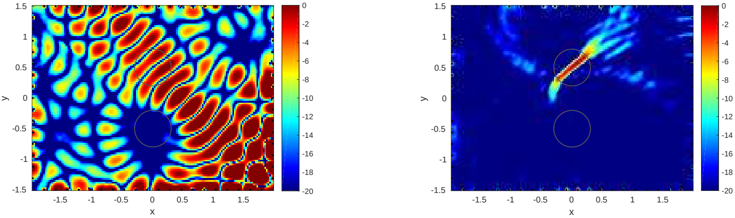
→ Reflections in 'dark zone' suppressed

Recent Developments in Reproduction IV

Joint Pressure and Velocity Matching (JPVM) for Multi-zone Reproduction (*Buerger, 2015, 2018*)

- $L = 70$ loudspeakers; image model order 3, reflection coefficient 0.9
- JPVM: 2 circular arrays with $N = 15$ microphones, radial distance 25mm;

Monofrequent sound field: Broadband soundfield ('impulse wave'):



$\rightarrow$ **JPVM $1 \ldots 7$ dB better than Pressure Matching**

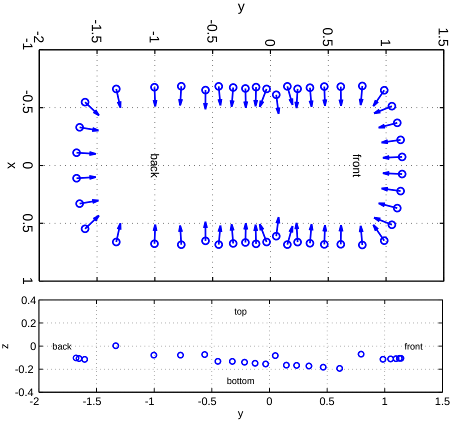
Some Challenges in Reproduction

Sparse, irregular, nonplanar loudspeaker arrays

Example: Loudspeaker array for a car

Top view:

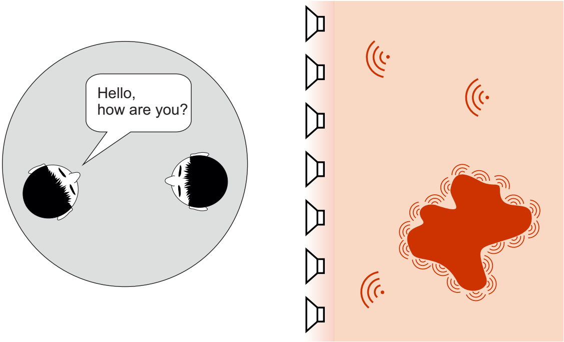
Side view:



- **Transforms for efficient sound field representation** (*Hofmann 2016*)

Some Challenges in Reproduction (cont'd)

Suppression of sound fields of spatially extended sources



- **Modelling of complex, partially incoherent sound fields**, e.g., noise through windows (*Buerger 2018)*

Generic Challenges in Reproduction

Adaptive Signal Processing:

- **Realistic time-varying scenarios**

  $\Rightarrow$ **Informed ... blind adaptive MIMO filtering for determining $\mathbf{H}_{\mathbf{wv}}$**

  $\Rightarrow$ **Learning of useful prior knowledge on listening scenarios**

  $\Rightarrow$ **Suppression of wideband nonstationary local noise and interference**

Psychoacoustics:

- **How much spatial realism do listeners require and/or appreciate?**

  – **when moving?**

  – **with other distracting sound sources?**

Generic Challenges in Reproduction

Adaptive Signal Processing:

- **Realistic time-varying scenarios**

  $\Rightarrow$ **Informed ... blind adaptive MIMO filtering for determining $\mathbf{H}_{\mathbf{wv}}$**

  $\Rightarrow$ **Learning of useful prior knowledge on listening scenarios**

  $\Rightarrow$ **Suppression of wideband nonstationary local noise and interference**

Psychoacoustics:

- **How much spatial realism do listeners require and/or appreciate?**

  – when moving?

  – with other distracting sound sources?

Overview

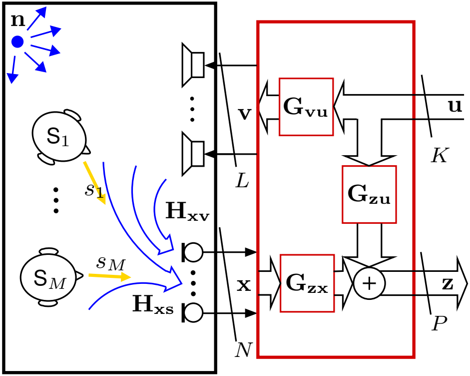
Reproduction

Acquisition

- MIMO acoustic echo cancellation
- Informed signal extraction
- Blind signal extraction
- Localization
- Interference and noise estimation
- Robot audition

Summary and Outlook

Fundamental Problems for Signal Acquisition



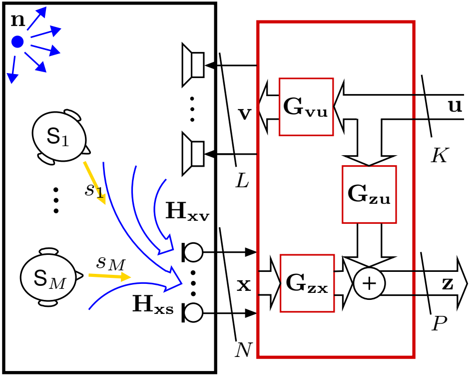
Goal: Undistorted source signals

$$\mathbf{z} = \mathbf{G_{zu}} * \mathbf{u} + \mathbf{G_{zx}} * \mathbf{x} \stackrel{!}{=} \mathbf{s} * \delta(k - k_0)$$

where $\mathbf{x} = \mathbf{H_{xs}} * \mathbf{s} + \mathbf{H_{xv}} * \mathbf{v} + \mathbf{n_x}$

- Echo cancellation:
  $(\mathbf{G_{zu}} + \mathbf{G_{zx}} * \mathbf{H_{xv}} * \mathbf{G_{vu}}) * \mathbf{u} = \mathbf{0}$
- Source separation and dereverberation:
  $\mathbf{G_{zx}} * \mathbf{H_{xs}} * \mathbf{s} = \mathbf{s} * \delta(k - k_0)$
- Noise and interference suppression:
  $\mathbf{G_{zx}} * \mathbf{n_x} = \mathbf{0}$

Fundamental Problems for Signal Acquisition



Goal: Undistorted source signals

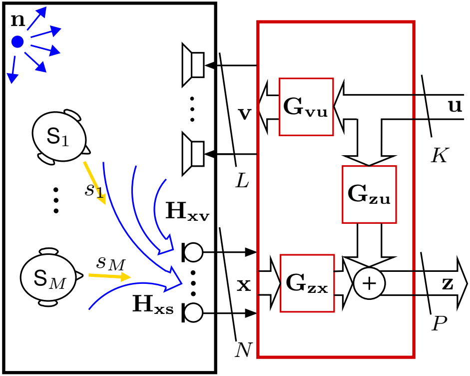
$$\mathbf{z} = \mathbf{G_{zu}} * \mathbf{u} + \mathbf{G_{zx}} * \mathbf{x} \stackrel{!}{=} \mathbf{s} * \delta(k - k_0)$$

where $\mathbf{x} = \mathbf{H_{xs}} * \mathbf{s} + \mathbf{H_{xv}} * \mathbf{v} + \mathbf{n_x}$

3 Subproblems:

- Echo cancellation:
  $$(\mathbf{G_{zu}} + \mathbf{G_{zx}} * \mathbf{H_{xv}} * \mathbf{G_{vu}}) * \mathbf{u} = \mathbf{0}$$
- Source separation and dereverberation:
  $$\mathbf{G_{zx}} * \mathbf{H_{xs}} * \mathbf{s} = \mathbf{s} * \delta(k - k_0)$$
- Noise and interference suppression:
  $$\mathbf{G_{zx}} * \mathbf{n_x} = \mathbf{0}$$

Fundamental Problems for Signal Acquisition



Goal: Undistorted source signals

$$\mathbf{z} = \mathbf{G_{zu}} * \mathbf{u} + \mathbf{G_{zx}} * \mathbf{x} \overset{!}{=} \mathbf{s} * \delta(k - k_0)$$

where $\mathbf{x} = \mathbf{H_{xs}} * \mathbf{s} + \mathbf{H_{xv}} * \mathbf{v} + \mathbf{n_x}$

3 Subproblems:

- Echo cancellation:

  $(\mathbf{G_{zu}} + \mathbf{G_{zx}} * \mathbf{H_{xv}} * \mathbf{G_{vu}}) * \mathbf{u} = 0$

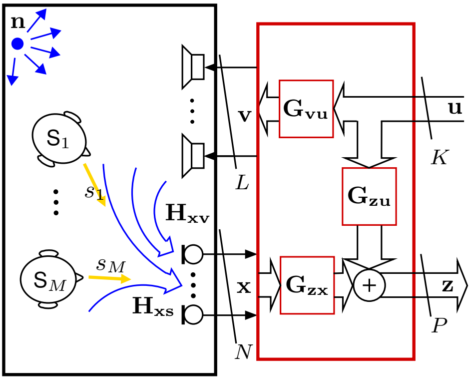
- Source separation and dereverberation:

  $\mathbf{G_{zx}} * \mathbf{H_{xs}} * \mathbf{s} = \mathbf{s} * \delta(k - k_0)$

- Noise and interference suppression:

  $\mathbf{G_{zx}} * \mathbf{n_x} = 0$

Fundamental Problems for Signal Acquisition



Goal: Undistorted source signals

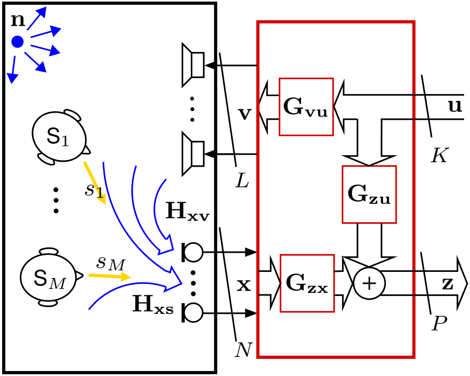
$$\mathbf{z} = \mathbf{G_{zu}} * \mathbf{u} + \mathbf{G_{zx}} * \mathbf{x} \overset{!}{=} \mathbf{s} * \delta(k - k_0)$$

where $\mathbf{x} = \mathbf{H_{xs}} * \mathbf{s} + \mathbf{H_{xv}} * \mathbf{v} + \mathbf{n_x}$

3 Subproblems:

- Echo cancellation:
  $$(\mathbf{G_{zu}} + \mathbf{G_{zx}} * \mathbf{H_{xv}} * \mathbf{G_{vu}}) * \mathbf{u} = \mathbf{0}$$
- Source separation and dereverberation:
  $$\mathbf{G_{zx}} * \mathbf{H_{xs}} * \mathbf{s} = \mathbf{s} * \delta(k - k_0)$$
- Noise and interference suppression:
  $$\mathbf{G_{zx}} * \mathbf{n_x} = \mathbf{0}$$

Fundamental Problems for Signal Acquisition



Goal: Undistorted source signals

$$\mathbf{z} = \mathbf{G_{zu}} * \mathbf{u} + \mathbf{G_{zx}} * \mathbf{x} \stackrel{!}{=} \mathbf{s} * \delta(k - k_0)$$

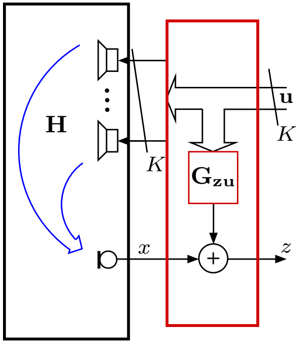
where $\mathbf{x} = \mathbf{H_{xs}} * \mathbf{s} + \mathbf{H_{xv}} * \mathbf{v} + \mathbf{n_x}$

3 Subproblems:

- Echo cancellation:
  $$(\mathbf{G_{zu}} + \mathbf{G_{zx}} * \mathbf{H_{xv}} * \mathbf{G_{vu}}) * \mathbf{u} = \mathbf{0}$$
- Source separation and dereverberation:
  $$\mathbf{G_{zx}} * \mathbf{H_{xs}} * \mathbf{s} = \mathbf{s} * \delta(k - k_0)$$
- Noise and interference suppression:
  $$\mathbf{G_{zx}} * \mathbf{n_x} = \mathbf{0}$$

Components of $\mathbf{x}$, i.e., $\mathbf{H_{xs}} * \mathbf{s}$, $\mathbf{H_{xv}} * \mathbf{v}$, $\mathbf{n_x}$, must be separated by $\mathbf{G}$!

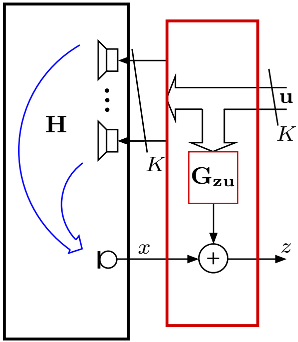
Multichannel AEC - Key Problems and State-of-the-Art



Key problems:

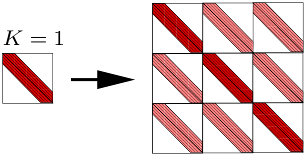
- **Computational complexity** - $\mathbf{G_{zu}} = [\mathbf{g}_1, \ldots, \mathbf{g}_K]$ with $L_G \gg 1000$ for each $\mathbf{g}_i$ to be continuously identified

Multichannel AEC - Key Problems and State-of-the-Art

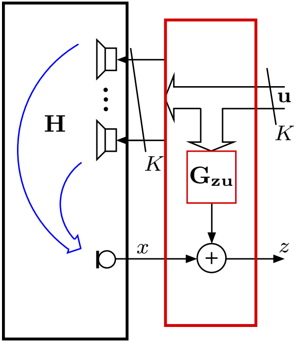


Key problems:

- **Computational complexity** -
  $\mathbf{G}_{\mathbf{zu}} = [\mathbf{g}_1, \ldots, \mathbf{g}_K]$ with $L_G \gg 1000$ for each $\mathbf{g}_i$ to be continuously identified
- **Correlation matrix $\mathrm{R}_{\mathbf{uu}}$ for Wiener solution**
  usually ill-conditioned, as $\mathbf{u}_i$ in $\mathbf{u} = [\mathbf{u}_1, \ldots, \mathbf{u}_K]$ mutually correlated

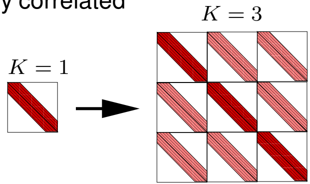


Multichannel AEC - Key Problems and State-of-the-Art



Key problems:

- **Computational complexity** - $\mathbf{G_{zu}} = [\mathbf{g}_1, \ldots, \mathbf{g}_K]$ with $L_G \gg 1000$ for each $\mathbf{g}_i$ to be continuously identified
- **Correlation matrix $\mathbf{R_{uu}}$ for Wiener solution** usually ill-conditioned, as $\mathbf{u}_i$ in $\mathbf{u} = [\mathbf{u}_1, \ldots, \mathbf{u}_K]$ mutually correlated



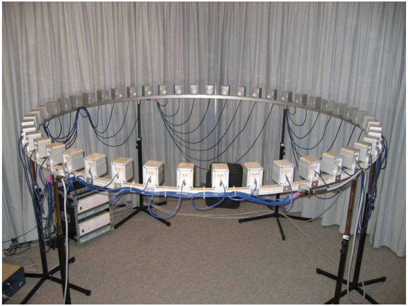
Example: Home-entertainment control
with $K = 5$, $K \cdot L_G > 20000$ (*Buchner 2003ff*)

Microphone x: Output $z = \hat{s}$:

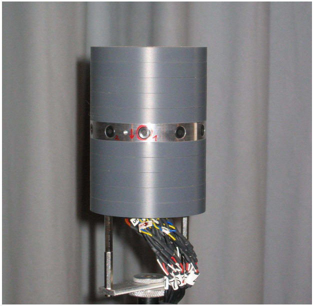
AEC for 'Massive' Multichannel Reproduction

Consider: Immersive environment with

$L = 48$ loudspeakers



$N = 10$ microphones

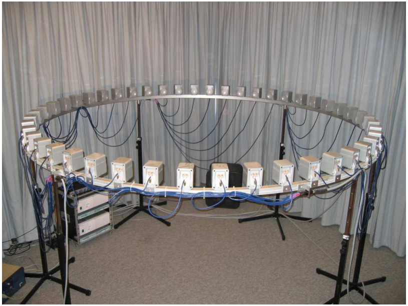


$\Longrightarrow L \times N = 480$ (equally important) adaptive echo cancellers!

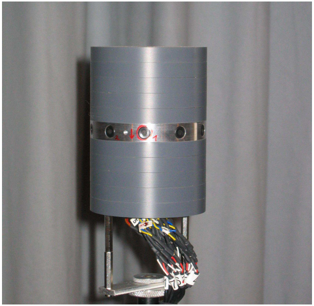
AEC for 'Massive' Multichannel Reproduction

Consider: Immersive environment with

$L = 48$ loudspeakers



$N = 10$ microphones



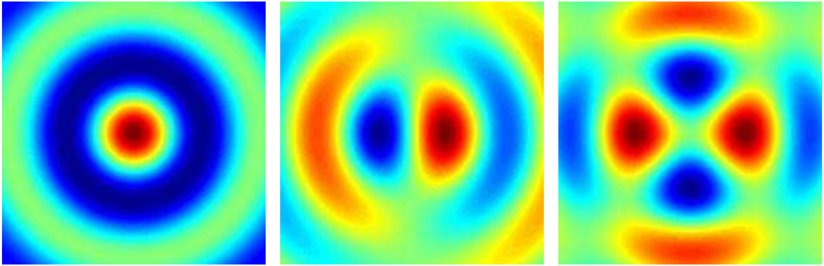
$\Longrightarrow L \times N = 480$ (equally important) adaptive echo cancellers!

Alternatives?

AEC in the Wave Domain

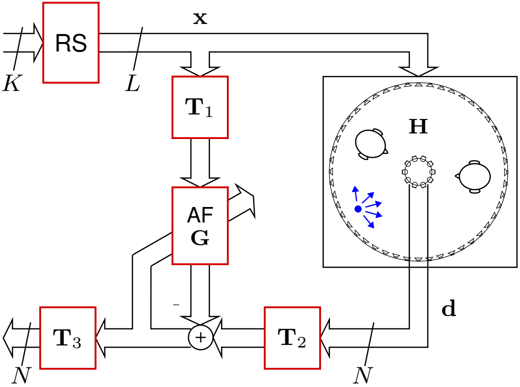
Idea: Represent the sound field by **spatial basis functions**, e.g., cylindrical harmonics in 2D /spherical harmonics in 3D

Illustration: Cylindrical Harmonics of order M=0,1,2 ($\omega = 2\pi 400$Hz, 1m$\times$1m)



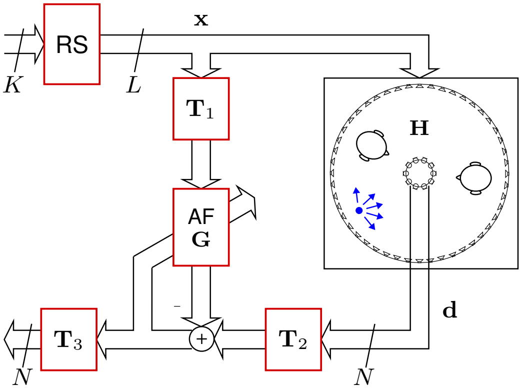
AEC in the Wave Domain

(*Buchner 2004; Schneider 2009ff*)



AEC in the Wave Domain

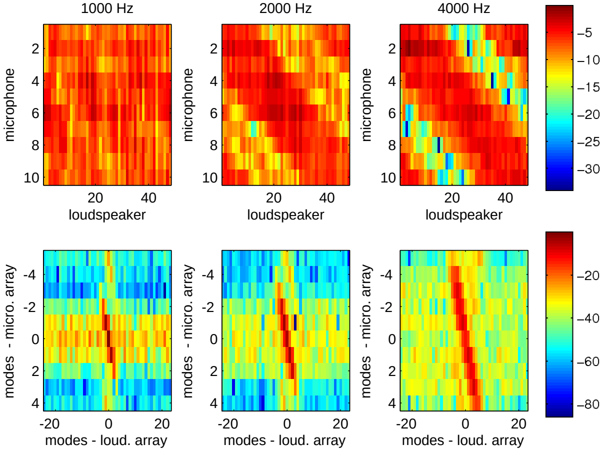
(*Buchner 2004; Schneider 2009ff*)



Promise: Echo cancellers only for dominant modes

AEC in the Wave domain (cont'd)

Spectral magnitudes of transfer functions for real room ($T_{60} \approx 300$ms):



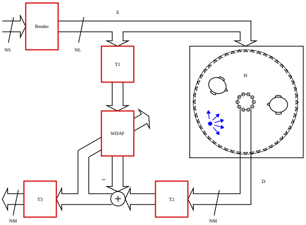
T/F domain

Wave domain

$\Longrightarrow$ **Dominant Modes in Wave domain**

AEC in the Wave domain (cont'd)

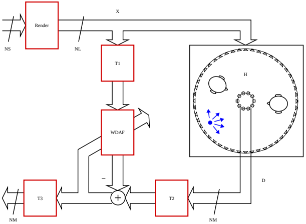
Real-time implementation on PC:



- $6 \ll 480$ **AECs** with
- $L_G = 8192$ coefficients each,
- adaptation by FDAF;
- $f_S = 22.05$kHz,
- $T_{60} \approx 300$ms;
- only 2D-wavefield modelled

AEC in the Wave domain (cont'd)

Real-time implementation on PC:

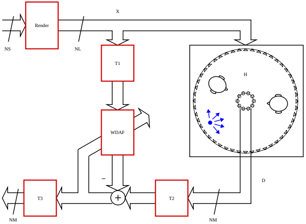


- $6 \ll 480$ **AECs** with
- $L_G = 8192$ coefficients each,
- adaptation by FDAF;
- $f_S = 22.05$kHz,
- $T_{60} \approx 300$ms;
- only 2D-wavefield modelled

$\Longrightarrow$ **Echo suppression (ERLE)** > 25**dB!**

AEC in the Wave domain (cont'd)

Real-time implementation on PC:



- $6 \ll 480$ **AECs** with
- $L_G = 8192$ coefficients each,
- adaptation by FDAF;
- $f_S = 22.05$kHz,
- $T_{60} \approx 300$ms;
- only 2D-wavefield modelled

$\Longrightarrow$ **Echo suppression (ERLE)** > 25**dB!**

Current Challenges:

- **General 3D array topologies** (Hofmann 2016)
- **Faster convergence via source-specific system identification** (Hofmann 2016)

Overview

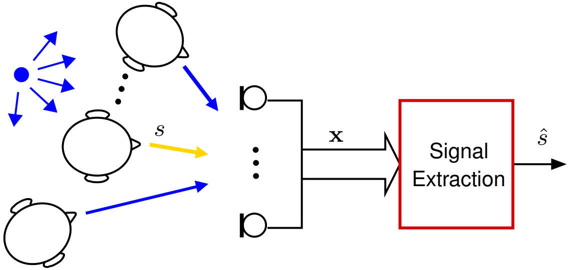
Reproduction

Acquisition

- MIMO acoustic echo cancellation
- Informed signal extraction
- Blind signal extraction
- Localization
- Interference and noise estimation
- Robot audition

Summary and Outlook

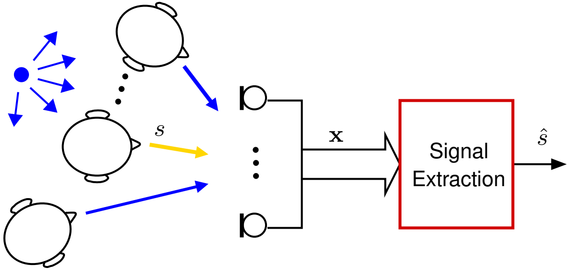
Acoustic Signal Extraction



Goal: Extract desired source ('target') from acoustic scene with

- nonstationary, nonwhite noise of unknown spatial coherence
- unknown number of nonstationary interfering point sources
- reverberation

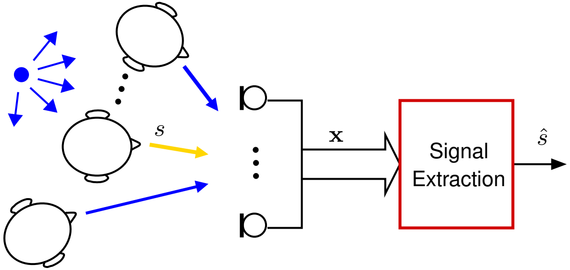
Acoustic Signal Extraction



Goal: Extract desired source ('target') from acoustic scene with

- nonstationary, nonwhite noise of unknown spatial coherence
- unknown number of nonstationary interfering point sources
- reverberation

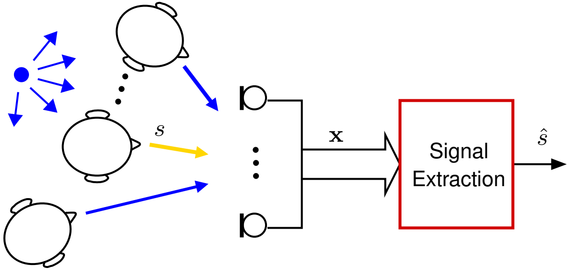
Acoustic Signal Extraction



Goal: Extract desired source ('target') from acoustic scene with

- nonstationary, nonwhite noise of unknown spatial coherence
- unknown number of nonstationary interfering point sources
- reverberation

Acoustic Signal Extraction



Goal: Extract desired source ('target') from acoustic scene with

- nonstationary, nonwhite noise of unknown spatial coherence
- unknown number of nonstationary interfering point sources
- reverberation

Multichannel Signal Extraction - Concepts

- **Informed Signal Extraction**, requiring, e.g, source localization
  - Data-independent spatiotemporal filtering, e.g., fixed beamformers, polynomial beamformers
  - Data-dependent, statistically optimum spatiotemporal filtering, e.g., Wiener Filter, MVDR/LCMV/LCMMI Beamformer, GSC
  - Learning-based optimum filtering, e.g., multichannel nonnegative matrix factorization (MC-NMF), dictionary-based (KSVD, PO-KSVD), ...

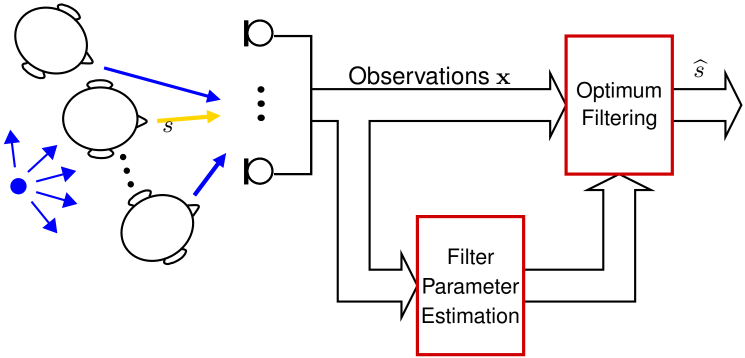
Multichannel Signal Extraction - Concepts

- **Informed Signal Extraction**, requiring, e.g, source localization
  - Data-independent spatiotemporal filtering, e.g., fixed beamformers, polynomial beamformers
  - Data-dependent, statistically optimum spatiotemporal filtering, e.g., Wiener Filter, MVDR/LCMV/LCMMI Beamformer, GSC
  - Learning-based optimum filtering, e.g., multichannel nonnegative matrix factorization (MC-NMF), dictionary-based (KSVD, PO-KSVD), ...
- **Blind Signal Extraction**
  - Subspace methods, PCA
  - Independent component/vector analysis (ICA/IVA/ILRMA), TRINICON

Multichannel Signal Extraction - Concepts

- **Informed Signal Extraction**, requiring, e.g, source localization
  - Data-independent spatiotemporal filtering, e.g., fixed beamformers, polynomial beamformers
  - Data-dependent, statistically optimum spatiotemporal filtering, e.g., Wiener Filter, MVDR/LCMV/LCMMI Beamformer, GSC
  - Learning-based optimum filtering, e.g., multichannel nonnegative matrix factorization (MC-NMF), dictionary-based (KSVD, PO-KSVD), . . .
- **Blind Signal Extraction**
  - Subspace methods, PCA
  - Independent component/vector analysis (ICA/IVA/ILRMA), TRINICON
- **Learning-based signal extraction** based on, e.g., HMM signal models; dictionaries; DNNs; manifold learning . . .

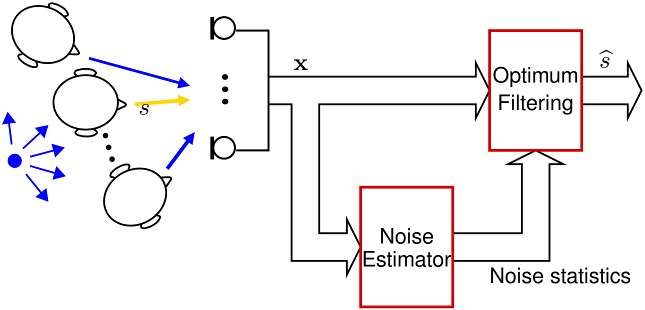
Informed Signal Extraction



- **Spatiotemporal Optimum Filtering** exploits diversity in time/frequency (T/F) and space
- **Filter Parameter Estimation**
  - can use statistical criteria (MMSE, ML, MAP, . . .) or (deep) learning
  - exploits prior knowledge

Informed Multichannel Signal Extraction (cont'd)

Concept 1: 'Optimum' Noise Suppression (MWF, SDW-MWF, MC-NMF, PO-KSVD, . . .)

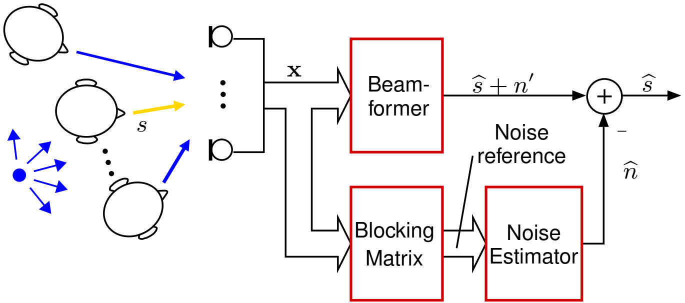


- exploits only magnitude and relative phase of noise/interference (via joint moments, learned multichannel dictionaries)
- will necessarily lead to some desired signal distortion

\+ can suppress any kind and number of noise/interference sources

Informed Multichannel Signal Extraction (cont'd)

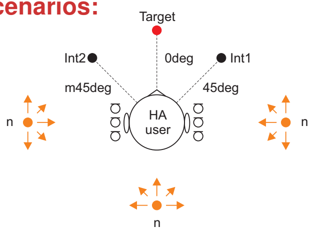
Concept 2: Interference & Noise Cancellation (LCMV, LCMMI, MVDR as GSC)

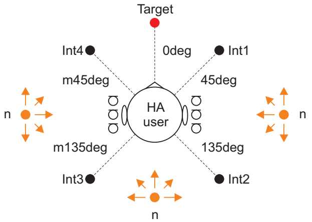


- \+ exploits magnitude and absolute phase of noise/interference
- \+ can perfectly suppress point sources (in free-field)
- \+ ideally, no target signal distortion
- \- limited performance for diffuse noise and reverberation

Informed Noise Suppression - Binaural Hearing Systems

Scenarios:





$d_{\text{mic}} = 15$cm, source-mic dist. 1.1m, $T_{60} \approx 300$ msec, $\mathbf{S}_{\mathbf{nn}}(\omega)$ **known!**

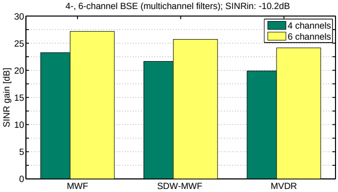
# Point sources	Input mon./bin.	# Micro-phones	MWF	SDW-MWF $(\lambda = 4)$	MVDR
3	/	4			
		6			
5	/	4			
		6			

Informed Multichannel Noise Suppression - Results (cont'd)

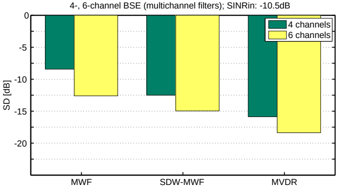
3 Point sources

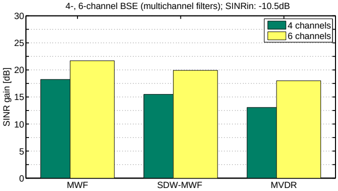
Target signal distortion (in dB)

Noise reduction gain (in dB)



5 Point sources





Multichannel Signal Extraction - Discussion

Results for known target location (DoA):

- Noise reduction >20 dB even in difficult scenarios for known $\mathbf{S_{nn}}(\omega)$
- Target signal distortion acceptable
- Suppression gains increase with # of microphones

Multichannel Signal Extraction - Discussion

Results for known target location (DoA):

- Noise reduction >20 dB even in difficult scenarios for known $\mathbf{S}_{\mathbf{nn}}(\omega)$
- Target signal distortion acceptable
- Suppression gains increase with # of microphones

Challenges

- Estimation of noise and interferer statistics, e.g., PSD matrix $\mathbf{S}_{\mathbf{nn}}(\omega)$
- Localization of target source
- Estimation of head-related or device-related relative transfer functions (HRTFs/DRTFs)

Multichannel Signal Extraction - Discussion

Results for known target location (DoA):

- Noise reduction >20 dB even in difficult scenarios for known $\mathbf{S_{nn}}(\omega)$
- Target signal distortion acceptable
- Suppression gains increase with # of microphones

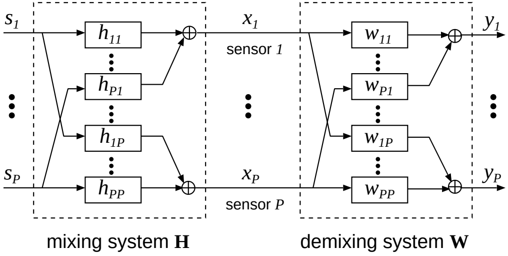
Challenges

- Estimation of noise and interferer statistics, e.g., PSD matrix $\mathbf{S_{nn}}(\omega)$
- Localization of target source
- Estimation of head-related or device-related relative transfer functions (HRTFs/DRTFs)

'Blind' methods as alternatives?

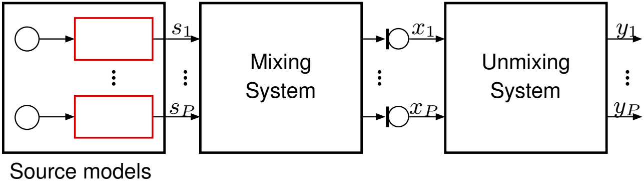
ICA-based Blind MIMO Signal Processing

Optimum Filtering by a linear MIMO demixing system $\mathbf{W}$

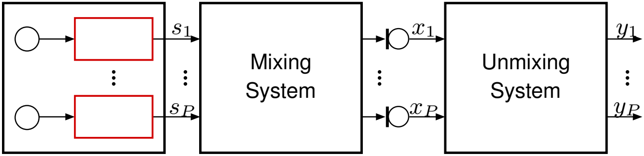


- demixing system is a matrix of adaptive FIR filters
- incoherent sound fields are considered as additive noise in each sensor
- interpretation as superposition of several MISO systems ($\rightarrow$ beamformers)
- ideal BSS corresponds to P ideal interference cancellers ($\rightarrow$ beamformers)

ICA-based Blind MIMO signal processing - TRINICON Concept



ICA-based Blind MIMO signal processing - TRINICON Concept



Generic TRINICON Cost Function (*Buchner et al. 2003ff*) exploits 3 N's (Nongaussianity, Nonstationarity, Nonwhiteness) for convolutive ICA

$$\mathcal{J}(m) = -\sum_{i=0}^{\infty} \beta(i,m) \frac{1}{N} \sum_{j=0}^{N-1} \{\log(\hat{p}_{s,PD}(\mathbf{y}(i,j))) - \log(\hat{p}_{y,PD}(\mathbf{y}(i,j)))\}$$

with PD-variate pdfs (P: source number, D: filter length)

- $\hat{p}_{s,PD}(\cdot)$ for source (assumed or estimated)
- $\hat{p}_{y,PD}(\cdot)$ for output

ICA-based Blind MIMO signal processing - TRINICON (cont'd)

Cost function for Second-order Statistics (SOS)

$$\mathcal{J}_{\mathrm{SOS}}(m) = \sum_{i=0}^{\infty} \beta(i,m) \left\{ \log \det \hat{\mathbf{R}}_{\mathbf{ss}}(i) - \log \det \hat{\mathbf{R}}_{\mathbf{yy}}(i) \right\}$$

with correlation matrices $\hat{\mathbf{R}}_{\mathbf{ss}}, \hat{\mathbf{R}}_{\mathbf{yy}}$ of size $PD \times PD$

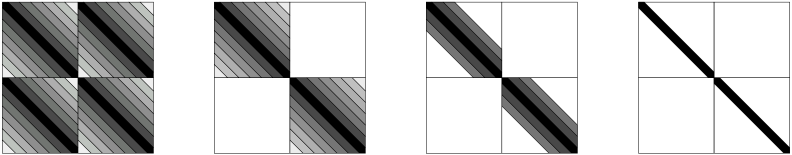
ICA-based Blind MIMO signal processing - TRINICON (cont'd)

Cost function for Second-order Statistics (SOS)

$$\mathcal{J}_{\mathrm{SOS}}(m) = \sum_{i=0}^{\infty} \beta(i,m) \left\{ \log \det \hat{\mathbf{R}}_{\mathbf{ss}}(i) - \log \det \hat{\mathbf{R}}_{\mathbf{yy}}(i) \right\}$$

with correlation matrices $\hat{\mathbf{R}}_{\mathbf{ss}}, \hat{\mathbf{R}}_{\mathbf{yy}}$ of size $PD \times PD$

Goals for SOS-TRINICON illustrated by $\hat{\mathbf{R}}_{\mathbf{ss}}$ (for $P = 2$ channels):



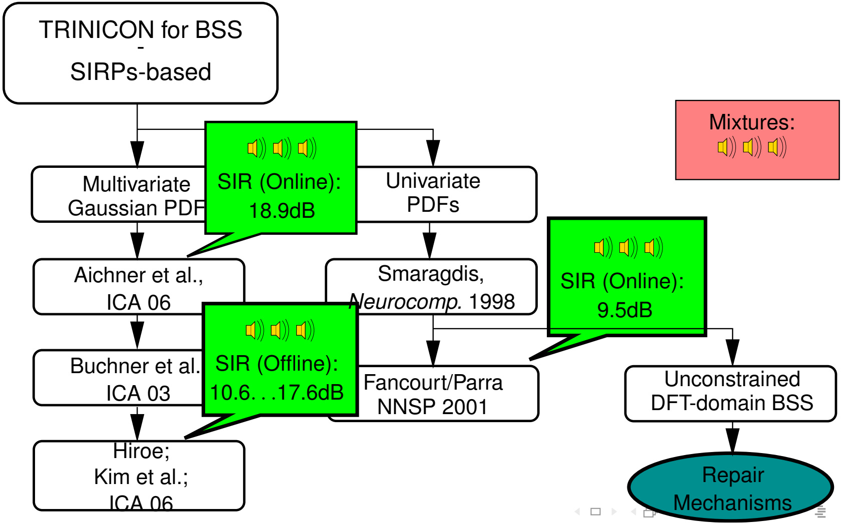
$\hat{\mathbf{R}}_{\mathbf{yy}} = \hat{\mathbf{R}}_{\mathbf{xx}}$
no processing

Source Separation (BSS)

Partial Deconvolution (**Dereverberation**)

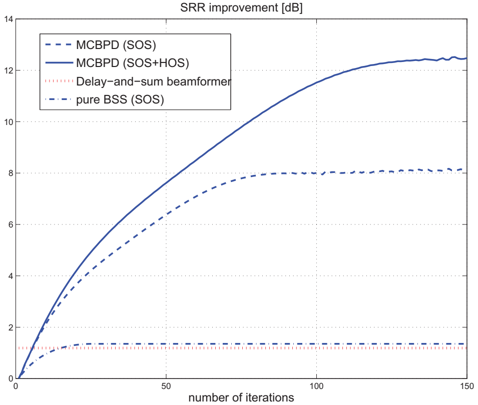
Deconvolution (MCBD)

Blind Spatial Filtering - TRINICON BSS Results



Blind Signal Separation & Dereverberation - TRINICON

Example: MCBPD dereverberation performance (*Buchner 2010*)



- $P = 2$ sources,
- $f_s = 16$kHz
- $Q = 4$ mics
- $d_{mic} = 16$cm,
- source-mic: 1.65m
- $T_{60} \approx 700$msec
- filter length 3000
- prediction order 32
- recording 30 sec
- offline adaptation

Sound examples: Mic: The outputs:

ICA-based BSS & Dereverberation - Discussion

- **Results**
  - BSS separation gain > 20dB for moderately reverberant rooms
  - Partial Deconvolution (MCBPD) reduces reverberation by up to 12dB

ICA-based BSS & Dereverberation - Discussion

- **Results**
  - BSS separation gain > 20dB for moderately reverberant rooms
  - Partial Deconvolution (MCBPD) reduces reverberation by up to 12dB
- **Conceptual Limitations**
  - external permutation can only be disambiguated by external knowledge
  - spatially white noise cannot be separated nor suppressed
  - number of point sources must be known

ICA-based BSS & Dereverberation - Discussion

- **Results**
  - BSS separation gain > 20dB for moderately reverberant rooms
  - Partial Deconvolution (MCBPD) reduces reverberation by up to 12dB
- **Conceptual Limitations**
  - external permutation can only be disambiguated by external knowledge
  - spatially white noise cannot be separated nor suppressed
  - number of point sources must be known
- **Challenges**
  - BSS algorithm for more sources, more reverberation
  - combining BSS with suppression of incoherent noise
  - fast and robust on-line algorithms for multisource dereverberation

Informed and Blind Signal Extraction - Synopsis

So far:

- Informed spatiotemporal filtering needs (at least) target location and noise statistics as reference
- Blind, ICA-based filtering cannot handle unconstrained scenarios

Informed and Blind Signal Extraction - Synopsis

So far:

- Informed spatiotemporal filtering needs (at least) target location and noise statistics as reference
- Blind, ICA-based filtering cannot handle unconstrained scenarios

Informed and Blind Signal Extraction - Synopsis

So far:

- Informed spatiotemporal filtering needs (at least) target location and noise statistics as reference
- Blind, ICA-based filtering cannot handle unconstrained scenarios

Informed and Blind Signal Extraction - Synopsis

So far:

- Informed spatiotemporal filtering needs (at least) target location and noise statistics as reference
- Blind, ICA-based filtering cannot handle unconstrained scenarios

Idea:

Use blind on-line algorithms and off-line learning algorithms to estimate reference information, such as,

- Source loci, DoAs

Informed and Blind Signal Extraction - Synopsis

So far:

- Informed spatiotemporal filtering needs (at least) target location and noise statistics as reference
- Blind, ICA-based filtering cannot handle unconstrained scenarios

Idea:

Use blind on-line algorithms and off-line learning algorithms to estimate reference information, such as,

- Source loci, DoAs
- Noise & reverberation statistics

Informed and Blind Signal Extraction - Synopsis

So far:

- Informed spatiotemporal filtering needs (at least) target location and noise statistics as reference
- Blind, ICA-based filtering cannot handle unconstrained scenarios

Idea:

Use blind on-line algorithms and off-line learning algorithms to estimate reference information, such as,

- Source loci, DoAs
- Noise & reverberation statistics
- Head/Device-related RTFs

Overview

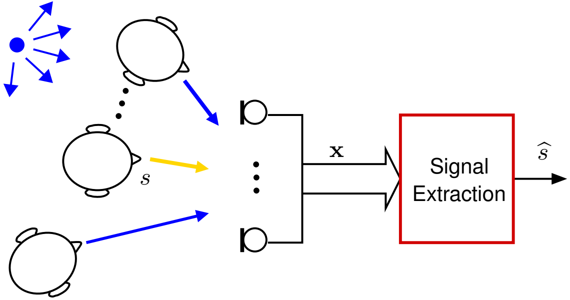
Reproduction

Acquisition

- MIMO acoustic echo cancellation
- Informed signal extraction
- Blind signal extraction
- Localization
- Interference and noise estimation
- Robot audition

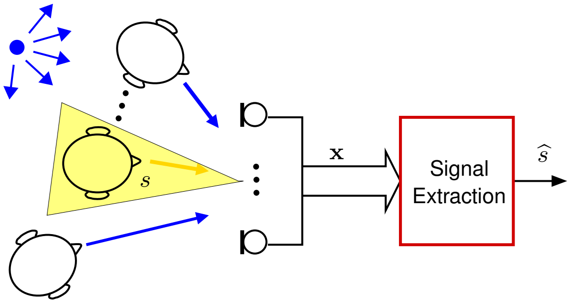
Summary and Outlook

ICA-based Localization for Signal Extraction



Why is localization difficult?

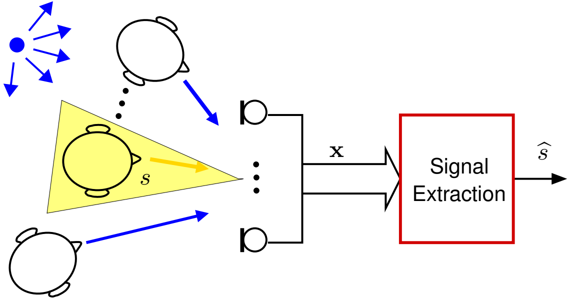
ICA-based Localization for Signal Extraction



Why is localization difficult?

- Multiple simultaneously active sources
- Multipath propagation due to reflections from walls

ICA-based Localization for Signal Extraction



Why is localization difficult?

- Multiple simultaneously active sources
- Multipath propagation due to reflections from walls

$\Rightarrow$ Not modeled by conventional methods (GCC, SRP, MUSIC, ESPRIT)

Source Localization - State of the Art

- Targets close to sensors, no strong early reflections
  - Single source: Correlation-based (GCC, GCC-PHAT)
  - Multiple sources: SRP-PHAT (*DiBiase 2001*), MUSIC/ESPRIT (*Teutsch 2005, Chen 2017*), ICA/IVA clustering (*Otsuka 2012*)

Source Localization - State of the Art

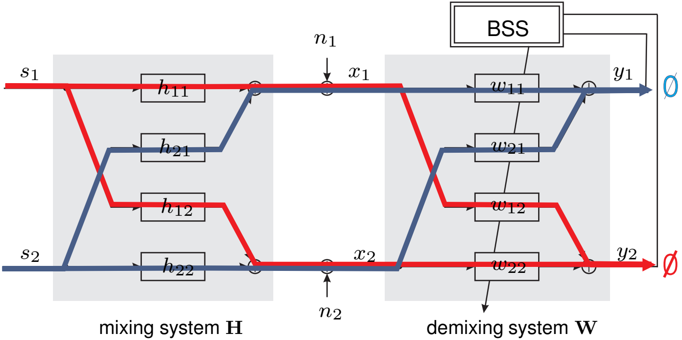
- Targets close to sensors, no strong early reflections
  - Single source: Correlation-based (GCC, GCC-PHAT)
  - Multiple sources: SRP-PHAT (*DiBiase 2001*), MUSIC/ESPRIT (*Teutsch 2005, Chen 2017*), ICA/IVA clustering (*Otsuka 2012*)
- Targets close to sensors, possibly strong early reflections
  - Blind RTF-estimation (BSI) by **ICA/TRINICON-BSS** (*Buchner 2005, Nesta 2009, Lombard 2012*)

Source Localization - State of the Art

- Targets close to sensors, no strong early reflections
  - Single source: Correlation-based (GCC, GCC-PHAT)
  - Multiple sources: SRP-PHAT (*DiBiase 2001*), MUSIC/ESPRIT (*Teutsch 2005, Chen 2017*), ICA/IVA clustering (*Otsuka 2012*)
- Targets close to sensors, possibly strong early reflections
  - Blind RTF-estimation (BSI) by **ICA/TRINICON-BSS** (*Buchner 2005, Nesta 2009, Lombard 2012*)
- Targets far from sensors
  - SRP-PHAT (strong reflections call for learned peak-picking strategies)
  - **ICA-based Averaged Directivity Pattern (ADP)** (*Lombard 2009ff*)
  - Blind RTF-estimation (BSI) by TRINICON-MCPD, see above (*Buchner 2005, 2010*)
  - Range estimation by **Coherent-to-Diffuse power Ratio (CDR)** (*Brendel 2018*)

BSI-based TDOA Estimation using ICA

2 Sources: BSS as interference canceller for 2-channel BSI (Reindl 2014)



Interference cancellation conditions $\rightarrow$ relative impulse responses

$$h_{11} * w_{12} + h_{12} * w_{22} = 0 \Rightarrow w_{12} = -h_{12} * h_{11}^{-1} * w_{22}$$

$$h_{21} * w_{11} + h_{22} * w_{21} = 0 \Rightarrow w_{21} = -h_{21} * h_{22}^{-1} * w_{11}$$

BSI-based TDOA Estimation using ICA (cont'd)

TDOA Estimation using ICA requires only the detection of direct acoustic paths in mixing system $\mathbf{H}$ (Buchner et al. 2005)

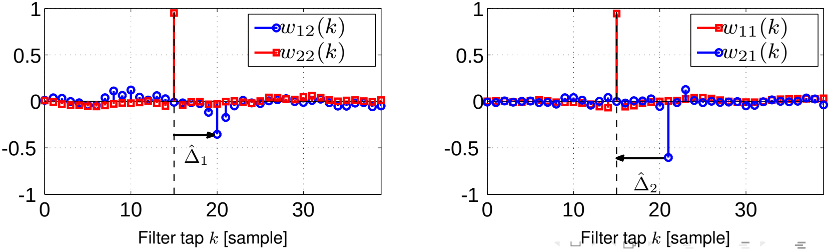
BSI-based TDOA Estimation using ICA (cont'd)

TDOA Estimation using ICA requires only the detection of direct acoustic paths in mixing system $\mathbf{H}$ (Buchner et al. 2005)

$\Longrightarrow$ **Compute TDOAs** from time difference between dominant peaks in w_{ij}:

$$\hat{\Delta}_1 = \underset{0\le k\le L-1}{\operatorname{argmax}} |w_{12,k}| - \underset{0\le k\le L-1}{\operatorname{argmax}} |w_{22,k}|$$

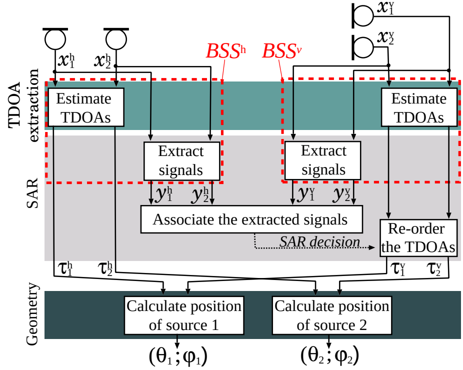
$$\hat{\Delta}_2 = \underset{0\le k\le L-1}{\operatorname{argmax}} |w_{11,k}| - \underset{0\le k\le L-1}{\operatorname{argmax}} |w_{21,k}|$$

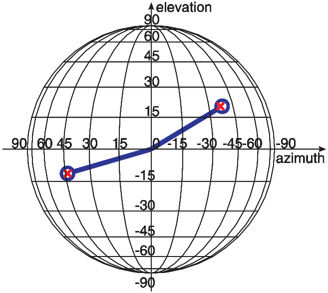


ICA-based TDOA Estimation for Multiple Sources in Multiple Dimensions

Localization ambiguity resolution by BSS (Lombard et al. 2006):

- BSS in each dimension provides separated outputs $\Longrightarrow$
- **Pairwise correlation BSS outputs in each dimension** resolves ambiguity





DOA Estimation using ADP

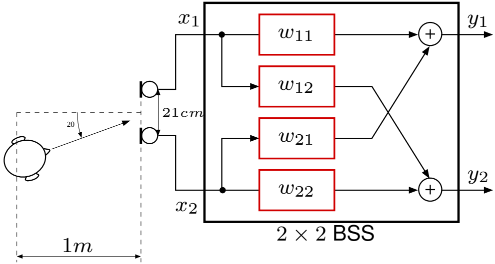
Averaged Directivity Patterns (ADP) for more than 2 sources

Example: 2×2- TRINICON-BSS (SOS), $T_{60} = 250$ms (Lombard et al., 2009)

DOA Estimation using ADP

Averaged Directivity Patterns (ADP) for more than 2 sources

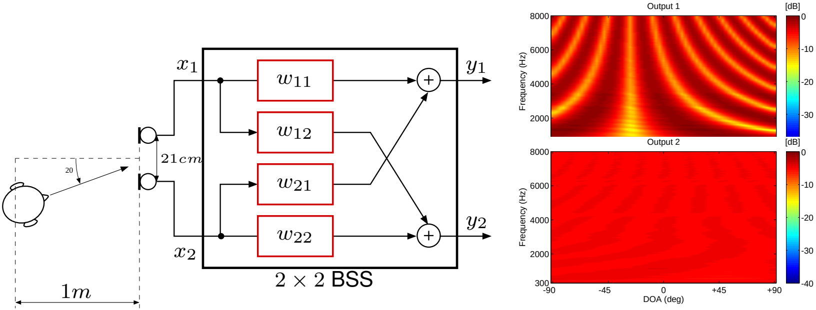
Example: 2×2- TRINICON-BSS (SOS), $T_{60} = 250$ms (Lombard et al., 2009)



DOA Estimation using ADP

Averaged Directivity Patterns (ADP) for more than 2 sources

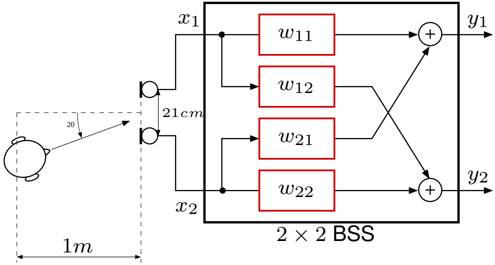
Example: 2×2- TRINICON-BSS (SOS), $T_{60} = 250$ms (Lombard et al., 2009)

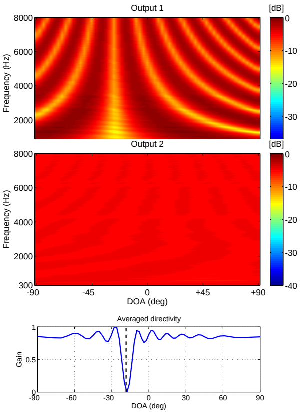


DOA Estimation using ADP

Averaged Directivity Patterns (ADP) for more than 2 sources

Example: 2×2- TRINICON-BSS (SOS), $T_{60} = 250$ms (Lombard et al., 2009)



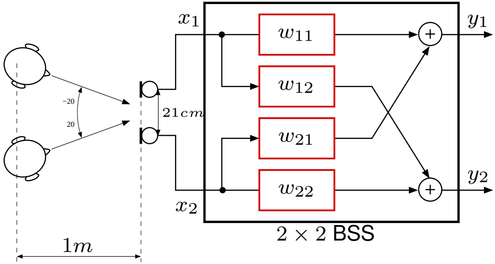


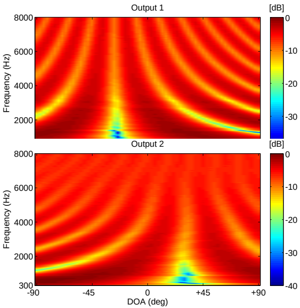
BSS doesn't need two active sources $\Leftarrow$

DOA Estimation using ADP

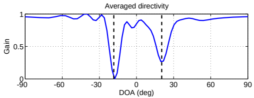
Averaged Directivity Patterns (ADP) for more than 2 sources

Example: 2×2- TRINICON-BSS (SOS), $T_{60} = 250$ms (Lombard et al., 2009)





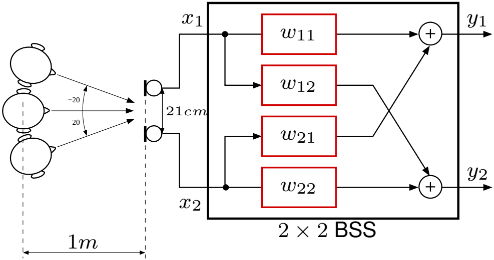
Regular case $\Longleftarrow$

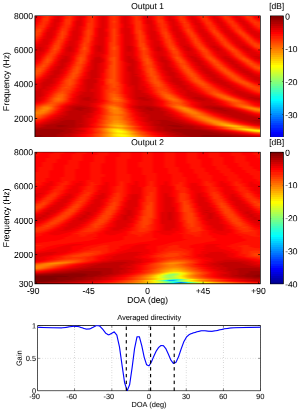


DOA Estimation using ADP

Averaged Directivity Patterns (ADP) for more than 2 sources

Example: 2×2- TRINICON-BSS (SOS), $T_{60} = 250$ms (Lombard et al., 2009)



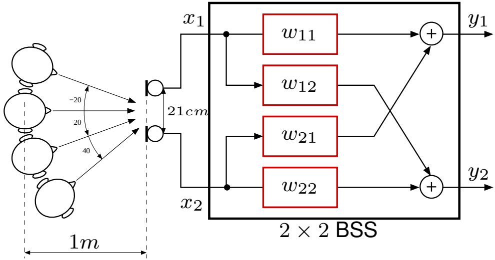


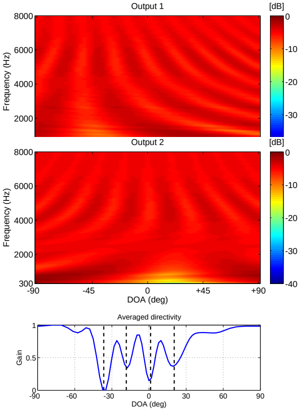
Still pronounced minima ⟸

DOA Estimation using ADP

Averaged Directivity Patterns (ADP) for more than 2 sources

Example: 2×2- TRINICON-BSS (SOS), $T_{60} = 250$ms (Lombard et al., 2009)



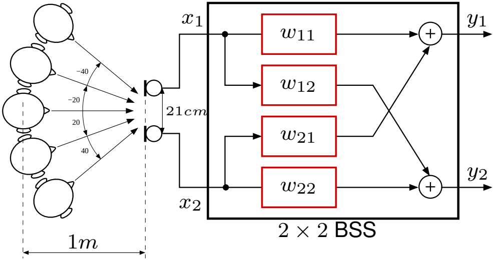


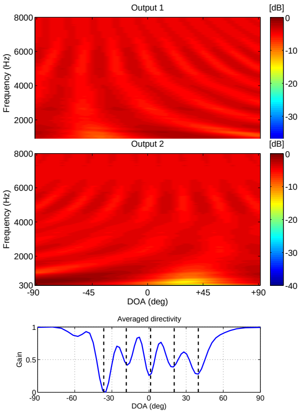
More minima ⟸

DOA Estimation using ADP

Averaged Directivity Patterns (ADP) for more than 2 sources

Example: 2×2- TRINICON-BSS (SOS), $T_{60} = 250$ms (Lombard et al., 2009)





5 sources localizable with 2 mics! ⟸

Range Estimation using Coherent-to-Diffuse Ratio (CDR)

CDR and Diffuseness ζ estimated from observed coherence Γ_{x} with assumed reverberation coherence Γ_{n} (*Schwarz 2015*)

$$\text{CDR} = \frac{\Gamma_{\mathrm{n}} - \Gamma_{\mathrm{x}}}{\Gamma_{\mathrm{x}} - \Gamma_{\mathrm{c}}}, \qquad \zeta = \frac{1}{1 + \text{CDR}}$$

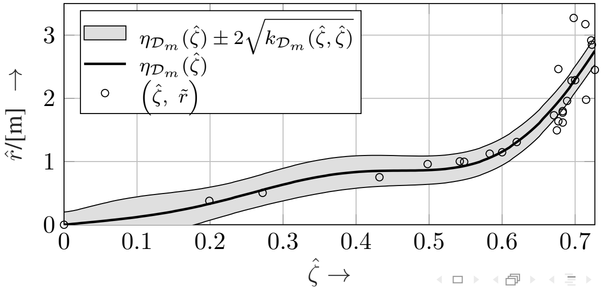
- **Environment-specific relation between range r and ζ** modelled by Gaussian Process (*Brendel 2018*)

Range Estimation using Coherent-to-Diffuse Ratio (CDR)

CDR and Diffuseness ζ estimated from observed coherence Γ_x with assumed reverberation coherence Γ_n (*Schwarz 2015*)

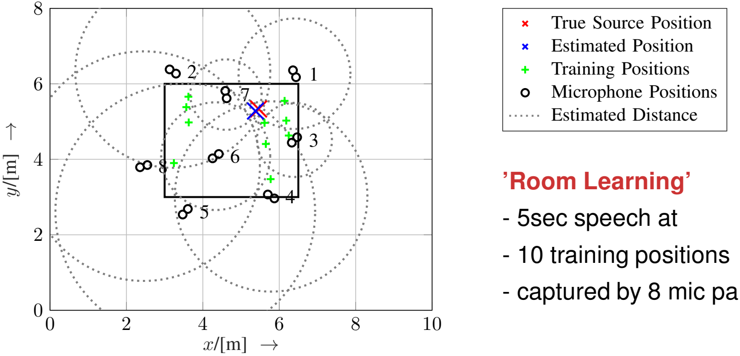
$$\mathsf{CDR} = \frac{\Gamma_\mathrm{n} - \Gamma_\mathrm{x}}{\Gamma_\mathrm{x} - \Gamma_\mathrm{c}}, \qquad \zeta = \frac{1}{1 + \mathsf{CDR}}$$

- **Environment-specific relation between range r and ζ** modelled by Gaussian Process (*Brendel 2018*)



Range Estimation using Coherent-to-Diffuse Ratio (CDR) (cont'd)

Example: Acoustic sensor network in 80m^2 room



'Room Learning'

- 5sec speech at
- 10 training positions
- captured by 8 mic pairs

Algorithm for learning and position estimation uniformly distributed over all network nodes (*Brendel 2018*)

Some Challenges in Localization and Tracking

Fusion of multiple instantaneous estimates for DoA, TDoA, CDR, . . . for non-stationary signals and time-varying scenarios accounting for

- uncertainty of individual estimates
- sensor utility in sensor networks

Tracking of multiple intermittently active targets

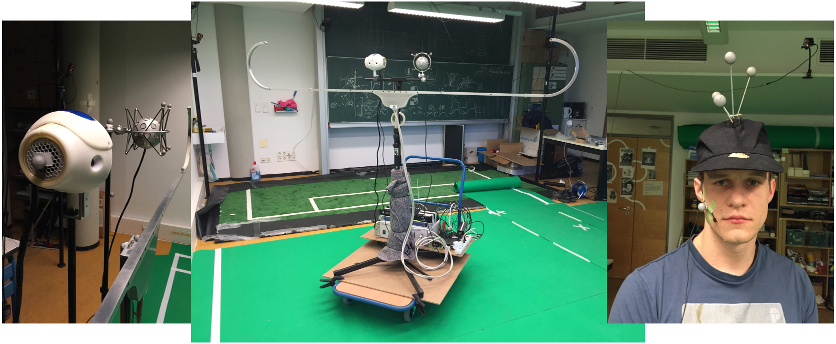
- Bayesian models, e.g., Extended Target Probability Hypothesis Density(PHD) Filter (*Brendel 2018* here!)
- Joint optimization with other modalities, esp., video (*Deleforge2015*)

$\Longrightarrow$ **4 Special Sessions at SAM 2018**

A Formal Challenge: IEEE AASP Challenge LOCATA

Tasks: Localization and tracking of

- one or more fixed or moving speech sources using
- 4 different fixed or moving microphone arrays



$\Longrightarrow$ **www.locata-challenge.org** and **SAM 2018, Special Session T1 L3**

Overview

Reproduction

Acquisition

- MIMO acoustic echo cancellation
- Informed signal extraction
- Blind signal extraction
- Localization
- Interference and noise estimation
- Robot audition

Summary and Outlook

Off-line estimation of Noise, Interference, and Reverberation

Off-line Noise Estimation ('learning', 'training')

- STFT-plane image processing (*Hofmann 2010*), e.g., for wind noise
- Multichannel NMF (*Ozerov 2010*)
- **PO-KSVD** (*Deleforge 2015*) for structured transient noise
- DNNs (*Xu 2015, Zhao 2016, Hershey 2016, Wu 2017*)

Off-line estimation of Noise, Interference, and Reverberation

Off-line Noise Estimation ('learning', 'training')

- STFT-plane image processing (*Hofmann 2010*), e.g., for wind noise
- Multichannel NMF (*Ozerov 2010*)
- **PO-KSVD** (*Deleforge 2015*) for structured transient noise
- DNNs (*Xu 2015, Zhao 2016, Hershey 2016, Wu 2017*)

Challenges:

- Robustness to unseen scenarios (e.g., unknown acoustic environments, interference types)
- Exploitation of spatial information from multichannel ('multistream') data by DNNs

On-line Estimation of Noise and Interference

- **Approach 1: Sparsity assumption.** Noise is estimated at instants in space-time-frequency where noise alone can be observed.
  - 3D space: Sparsity assured for point sources in freefield, reduced by reflections and scattering
  - T/F plane: Sparsity reduced by reverberation and increasing # of sources

On-line Estimation of Noise and Interference

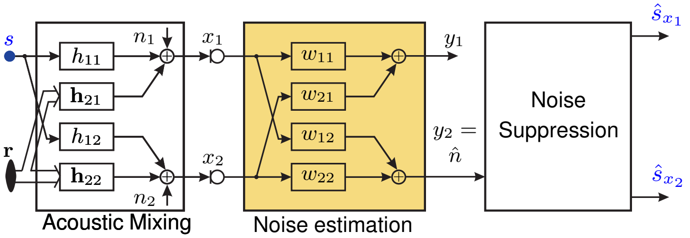
- **Approach 1: Sparsity assumption.** Noise is estimated at instants in space-time-frequency where noise alone can be observed.
  - 3D space: Sparsity assured for point sources in freefield, reduced by reflections and scattering
  - T/F plane: Sparsity reduced by reverberation and increasing # of sources
- Example: Noise statistics $\mathbf{S}_{\mathbf{nn}}$ for MWF, LCMV, MVDR estimated during target inactivity. Target activity/dominance detection uses, e.g,
  - Multichannel features: Coherence, crosscorrelation, DoA-dependent SNR estimates
  - Bayesian estimators with trained thresholds (*Taseska 2015*)
  - ANNs/DNNs after training (*Meier 2016f*)

On-line Estimation of Noise and Interference

- **Approach 1: Sparsity assumption.** Noise is estimated at instants in space-time-frequency where noise alone can be observed.
  - 3D space: Sparsity assured for point sources in freefield, reduced by reflections and scattering
  - T/F plane: Sparsity reduced by reverberation and increasing # of sources
- Example: Noise statistics $\mathbf{S_{nn}}$ for MWF, LCMV, MVDR estimated during target inactivity. Target activity/dominance detection uses, e.g,
  - Multichannel features: Coherence, crosscorrelation, DoA-dependent SNR estimates
  - Bayesian estimators with trained thresholds (*Taseska 2015*)
  - ANNs/DNNs after training (*Meier 2016f*)
- **Approach 2: Signal Separation** separating target from all noise and interference $\rightarrow$ **'Constrained ICA'**

ICA-based On-line Estimation of Noise and Interference

Signal model:



Idea: Extend ICA-BSS cost function by geometric constraint suppressing signals with target TDOA τ_s ('null steering'):

$$\mathcal{J}_{\mathrm{C-ICA}} = \widehat{\mathcal{E}}\left\{\log\left[\frac{\hat{p}_{Y_1Y_2}(\mathbf{y}_1,\mathbf{y}_2)}{\hat{p}_{Y_1}(\mathbf{y}_1)\hat{p}_{Y_2}(\mathbf{y}_2)}\right]\right\} + \eta\,\|w_{12}(k) + w_{22}(k-\tau_s)\|_2^2$$

Constrained ICA-based Estimation of Noise and Interference

Relation to Blocking Matrix of GSC for MVDR and LCMV implementations

- Null-steering constraint $\Longrightarrow$ blocking matrix constraint for MVDR

++ Constrained ICA continuously identifies RTFs to target

Constrained ICA-based Estimation of Noise and Interference

Relation to Blocking Matrix of GSC for MVDR and LCMV implementations

- Null-steering constraint $\Longrightarrow$ blocking matrix constraint for MVDR

++ Constrained ICA continuously identifies RTFs to target

++ Constrained ICA suppresses direct path + correlated reflections of target source

$\Rightarrow$ No 'target signal leakage'

$\Rightarrow$ Estimate for all undesired components (including diffuse noise and reverberation)

++ Unlike GSC blocking matrices, BSS is robust to target DoA errors

Constrained ICA-based Estimation of Noise and Interference

Relation to Blocking Matrix of GSC for MVDR and LCMV implementations

- Null-steering constraint $\Longrightarrow$ blocking matrix constraint for MVDR

++ Constrained ICA continuously identifies RTFs to target

++ Constrained ICA suppresses direct path + correlated reflections of target source

$\Rightarrow$ No 'target signal leakage'

$\Rightarrow$ Estimate for all undesired components (including diffuse noise and reverberation)

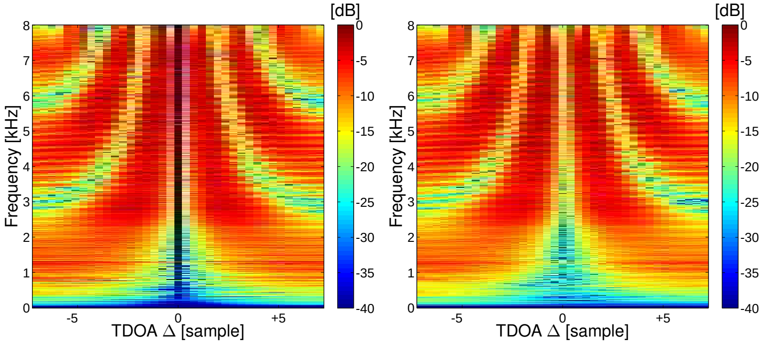
++ Unlike GSC blocking matrices, BSS is robust to target DoA errors

++ **Up-to-date noise estimate without source activity monitoring** (no voice activity detection or speech presence probability required)

ICA-based Blocking Matrix vs. Delay & Subtract Beamformer

Magnitude responses [dB] of mixing+demixing system ($\mathbf{C} = \mathbf{HW}$):

target at $\Delta = 0$, distance 1.1m; $d_{\text{mic}} = 15$cm, $T_{60} \approx 400$ms



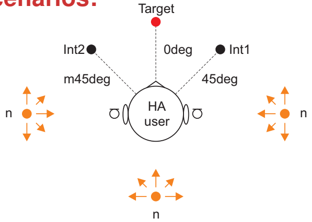
(a) C-ICA

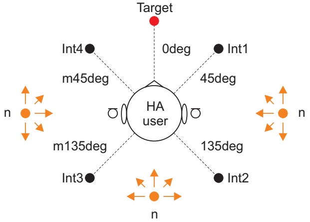
(b) Delay & Subtract BF

$\Longrightarrow$ Improved target suppression by C-ICA $\rightarrow$ Improved noise reference

Evaluation: Noise and Interference - Estimation and Suppression

Scenarios:





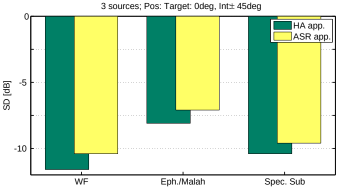
$d_{\text{mic}} = 15$cm, Source distance 1.1m, $T_{60} \approx 300$ msec, f_s=16kHz

# Point sources	Input (binaural)	Noise Estimate	Task	WF	Ephr./ Malah	Spec. Subtr.
3			Enh.			
			ASR			
5			Enh.			
			ASR			

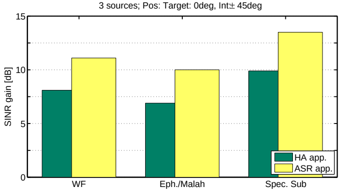
Evaluation: Noise and Interference Suppression (cont'd)

3 Point sources

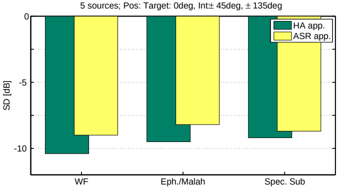
Signal Distortion (in dB)

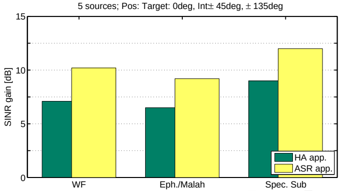


SINR Gain (in dB)



5 Point sources





Constrained ICA-based Noise Estimation - Generalizations

- **Estimation of diffuse reverberation** by suppression of the coherent target component (*Schwarz 2012; Reindl 2015*)

Constrained ICA-based Noise Estimation - Generalizations

- **Estimation of diffuse reverberation** by suppression of the coherent target component (*Schwarz 2012; Reindl 2015*)
- **Extraction of multiple point sources**. With given target DOAs
  - for all target sources, all RTFs can be simultaneously estimated by multiple 2-channel Constrained ICA units, producing
  - Blocking matrices for multiple parallel LCMV/LCMMI systems (*Reindl 2013ff, Markovich-Golan 2017*)

  Example: 4 Speech Sources (8 mics, SNR 30dB)

  Mic: One of 4 outputs:

Constrained ICA-based Noise Estimation - Generalizations

- **Estimation of diffuse reverberation** by suppression of the coherent target component (*Schwarz 2012; Reindl 2015*)
- **Extraction of multiple point sources**. With given target DOAs
  - for all target sources, all RTFs can be simultaneously estimated by multiple 2-channel Constrained ICA units, producing
  - Blocking matrices for multiple parallel LCMV/LCMMI systems (*Reindl 2013ff, Markovich-Golan 2017*)

  Example: 4 Speech Sources (8 mics, SNR 30dB)

  Mic: One of 4 outputs:

Challenge: Tracking of moving sources

Overview

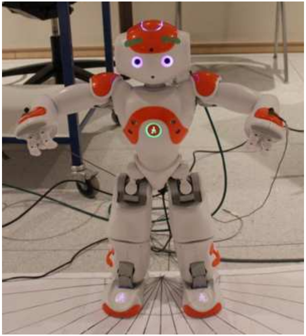
Reproduction

Acquisition

- MIMO acoustic echo cancellation
- Informed signal extraction
- Blind signal extraction
- Localization
- Interference and noise estimation
- Robot audition

Summary and Outlook

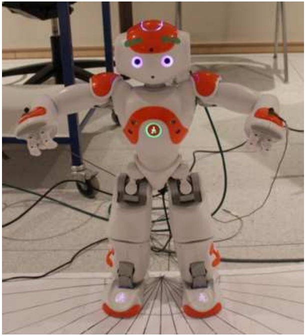
Another Challenge:
Multichannel Acoustic Signal Processing for Humanoïd Robots



Humanoïd Robots should . . .

- understand (multiple) human speech signals from a distance and in noise
- classify acoustic events to understand the acoustic scene

Another Challenge:
Multichannel Acoustic Signal Processing for Humanoïd Robots



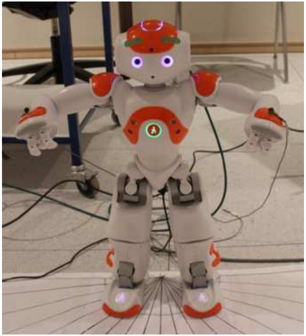
Humanoïd Robots should . . .

- understand (multiple) human speech signals from a distance and in noise
- classify acoustic events to understand the acoustic scene

This **requires**

- acoustic echo cancellation
- multiple source localization and tracking
- source activity monitoring (for ASR and attention model)
- multiple signal extraction, i.e., separation and dereverberation, for successful classification

Another Challenge:
Multichannel Acoustic Signal Processing for Humanoïd Robots



Humanoïd Robots should . . .

- understand (multiple) human speech signals from a distance and in noise
- classify acoustic events to understand the acoustic scene

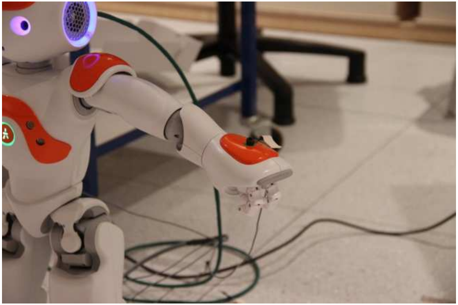
This **requires**

- acoustic echo cancellation
- multiple source localization and tracking
- source activity monitoring (for ASR and attention model)
- multiple signal extraction, i.e., separation and dereverberation, for successful classification

. . . and **offers New Options** and a **Robot-specific Problem: 'Ego Noise'**

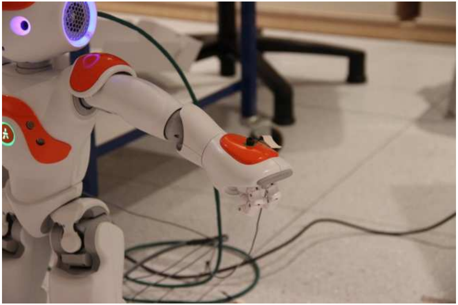
New options for microphone array topologies

Idea: Additional microphones at the limbs



New options for microphone array topologies

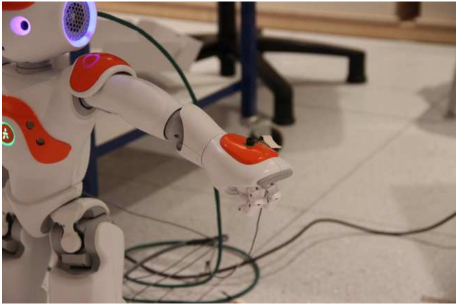
Idea: Additional microphones at the limbs



- **Promise:** Larger aperture, more distant from ego-noise sources

New options for microphone array topologies

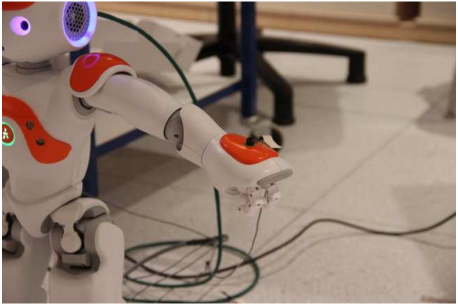
Idea: Additional microphones at the limbs



- **Promise:** Larger aperture, more distant from ego-noise sources
- **Challenge:** Spatial filtering for time-varying array topology

New options for microphone array topologies

Idea: Additional microphones at the limbs

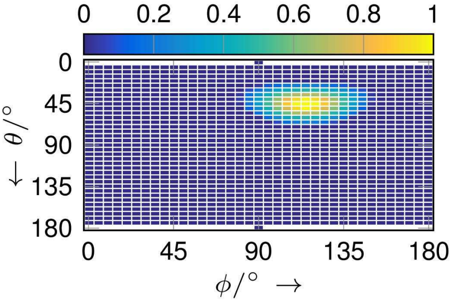


- **Promise:** Larger aperture, more distant from ego-noise sources
- **Challenge:** Spatial filtering for time-varying array topology
- **Approach:** ICA-based signal extraction with competing subsets of microphones (*Barfuss 2014*)

New options for microphone array topologies (cont'd)

Idea: Optimize microphone array topology for localization and signal extraction (*Tourbabin 2015*)

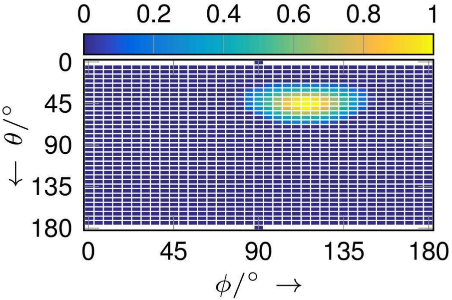
Example: HRTF-based fixed 2D polynomial beamforming using 12 mics (*Barfuss 2017*)



New options for microphone array topologies (cont'd)

Idea: Optimize microphone array topology for localization and signal extraction (*Tourbabin 2015*)

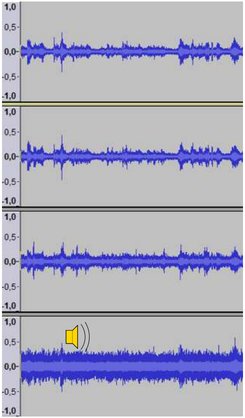
Example: HRTF-based fixed 2D polynomial beamforming using 12 mics (*Barfuss 2017*)



Challenges: Adaptive spatial filtering in combination with limb array

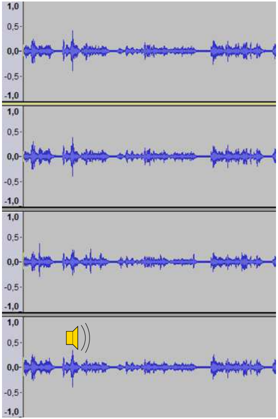
Robot Audition: 'Ego noise' from fan (Nao$^{(\mathrm{TM})}$)

4 head microphones:



Multichannel Wiener Filter $\Longrightarrow$

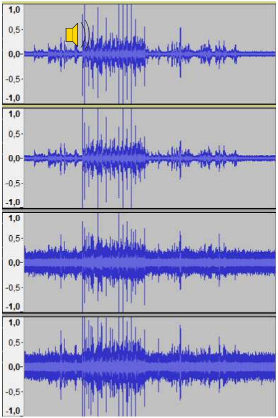
Output:



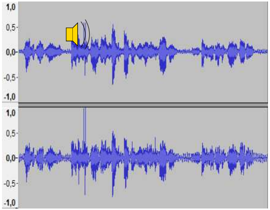
$\Longrightarrow$ **Stationary fan noise can be largely suppressed**

Robot Audition: 'Ego noise' from movements (Nao$^{(\mathrm{TM})}$)

Walking:

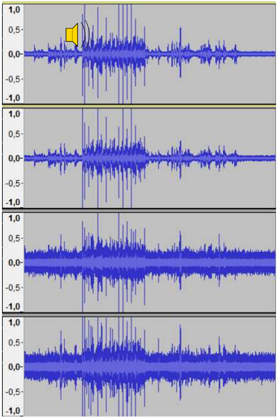


Head movement:

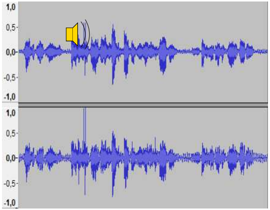


Robot Audition: 'Ego noise' from movements (Nao$^{(\mathrm{TM})}$)

Walking:



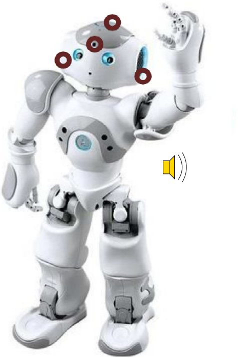
Head movement:



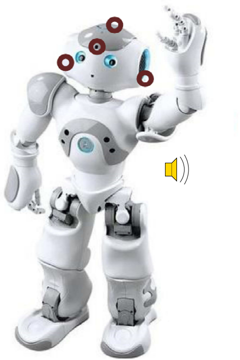
Challenge:

**Learn or model,
and suppress,
nonstationary noise**

Multichannel Ego Noise Estimation



Multichannel Ego Noise Estimation



Concept: (Deleforge 2015)

- **Training:** Learn noise characteristics from training data to create a (small) dictionary
- **Sparse Modeling:** Match current observations with few entries of dictionary ('atoms')
- **Optimum filtering:** Subtract few atoms (complex STFTs) with optimum phase from the noisy spectrum

A Sparse Model for Multichannel Mixtures

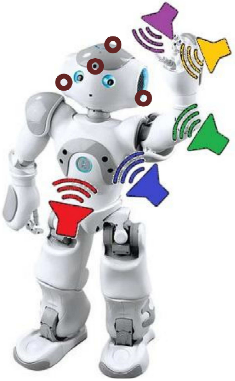
Proposed Model:

- M microphones

A Sparse Model for Multichannel Mixtures

Proposed Model:

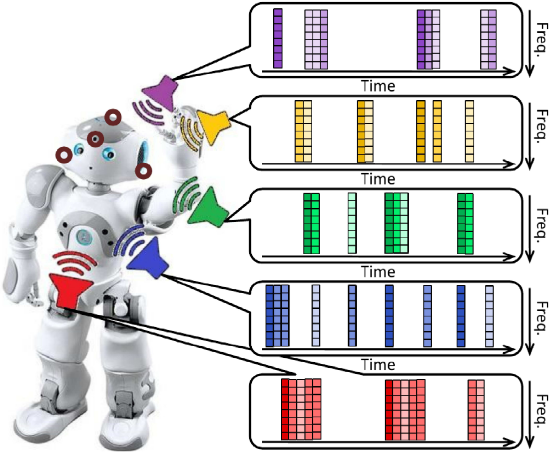
- M microphones
- $K \gg M$ sound sources



A Sparse Model for Multichannel Mixtures

Proposed Model:

- M microphones
- $K \gg M$ sound sources
  - Each source associated to a spectral component
  - Sparsely activated over time



A Sparse Mixture Model for Multichannel Mixtures

$\mathbf{Y} \in \mathbb{C}^{MF\times T}$ is an observed $M-$channel spectrogram with F frequency bins and T time frames

$$\mathbf{Y} = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T] = \left[\begin{pmatrix} \boldsymbol{y}_{11} \\ \boldsymbol{y}_{21} \\ \vdots \\ \boldsymbol{y}_{F1} \end{pmatrix}, \cdots, \begin{pmatrix} \boldsymbol{y}_{1T} \\ \boldsymbol{y}_{2T} \\ \vdots \\ \boldsymbol{y}_{FT} \end{pmatrix} \right] \text{ with } \boldsymbol{y}_{ij} = \begin{pmatrix} y_{ij,1} \\ \vdots \\ y_{ij,M} \end{pmatrix}$$

Each M-channel element of the T/F plane, $\boldsymbol{y}_{ft}$, is sparsely composed from K atoms:

$$\boldsymbol{y}_{ft} = \textstyle\sum_{k=1}^{K} \phi_{ft,k} p_{fk} \boldsymbol{a}_{fk} x_{kt} + \boldsymbol{e}_{ft},$$

p_{fk}: real-valued source power

$\boldsymbol{a}_{fk}$: complex-valued transfer function

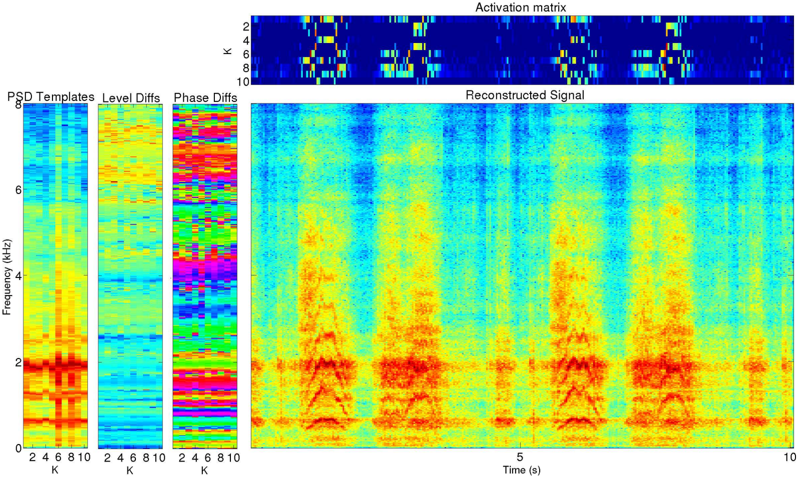
form dictionary entries $\boldsymbol{d}_{fk} = p_{fk}\boldsymbol{a}_{fk}$

x_{kt}: sparse real-valued activation gain

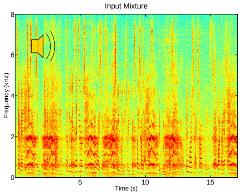
$\phi_{ft,k}$: **phase of atom** k

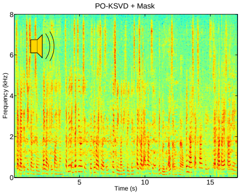
to be optimized!

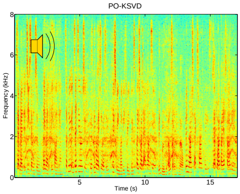
A Sparse Model for Multichannel Mixtures - Spectrogram Modeling



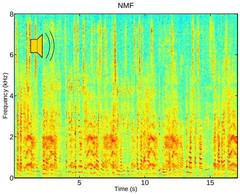
Phase-optimized K-SVD - Results for a Simple Movement

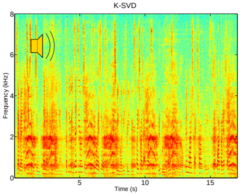






- $M = 4$ microphones
- **Dictionary:** $K = 10$
- **Training:** 1 min of robot arm waving + fan noise (suppressed by MWF)
- **Phase optimization** by Orthogonal Matching Pursuit (Deleforge 2015)

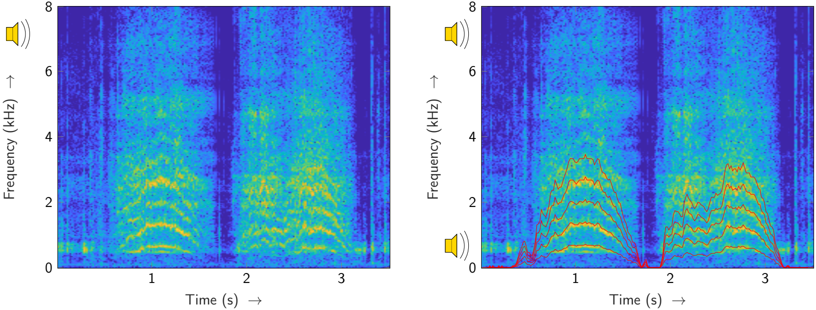




Ego-Noise Modeling and Learning for Complex Movements

Idea: Use motor data to predict harmonic noise components and avoid combinatorial increase of dictionary sizes (*Schmidt 2018*)

Example: Complex movement involving 6 joints with prediction of harmonics and dictionary size $K = 10$



Modeling of Dominant Components allows Small Dictionary

Robot Audition - Open Challenges

General

- Movements of robot and its microphone arrays
- Adaptation to changing acoustic environments
- Collaboration with other functions of the cognitive dynamic system

Robot Audition - Open Challenges

General

- Movements of robot and its microphone arrays
- Adaptation to changing acoustic environments
- Collaboration with other functions of the cognitive dynamic system

Specific

- Ego noise prediction for complex movements
- Exploitation of adaptable microphone array topology
- Joint optimization of signal extraction algorithms and robot gestures in cooperation with the robot's attention model

Summary

Signal processing challenges

- Complexity and time-variance of the acoustic environments
- Missing reference signals at the users - 'blindness'

Some examples

- Multizone Sound Reproduction
- MIMO Acoustic Echo Cancellation
- Informed and Blind Signal Extraction for Multiple Sources
- Obtaining Reference Information - Localization, Noise and Interference Estimation and Learning
- Robot Audition

Summary

Signal processing challenges

- Complexity and time-variance of the acoustic environments
- Missing reference signals at the users - 'blindness'

Some examples

- Multizone Sound Reproduction
- MIMO Acoustic Echo Cancellation
- Informed and Blind Signal Extraction for Multiple Sources
- Obtaining Reference Information - Localization, Noise and Interference Estimation and Learning
- Robot Audition

Concluding Remarks

Acoustic signal acquisition and reproduction challenged SP for 50+ years

⇒ many useful practical solutions

⇒ many longstanding and upcoming relevant problems still unsolved

Increasingly complex SP problems call for

- increasingly intricate and powerful, theoretically well-founded algorithms
- advanced physics-based modelling
- machine learning where physics-based models fail

Concluding Remarks

Acoustic signal acquisition and reproduction challenged SP for 50+ years

⇒ many useful practical solutions

⇒ many longstanding and upcoming relevant problems still unsolved

Increasingly complex SP problems call for

- increasingly intricate and powerful, theoretically well-founded algorithms
- advanced physics-based modelling
- machine learning where physics-based models fail

Concluding Remarks

Acoustic signal acquisition and reproduction challenged SP for 50+ years

⇒ many useful practical solutions

⇒ many longstanding and upcoming relevant problems still unsolved

Increasingly complex SP problems call for

- increasingly intricate and powerful, theoretically well-founded algorithms
- advanced physics-based modelling
- machine learning where physics-based models fail

We don't need to learn what we can model,

but we should learn what we can't model,

Concluding Remarks

Acoustic signal acquisition and reproduction challenged SP for 50+ years

⇒ many useful practical solutions

⇒ many longstanding and upcoming relevant problems still unsolved

Increasingly complex SP problems call for

- increasingly intricate and powerful, theoretically well-founded algorithms
- advanced physics-based modelling
- machine learning where physics-based models fail

We don't need to learn what we can model,

but we should learn what we can't model,

and find better models . . .