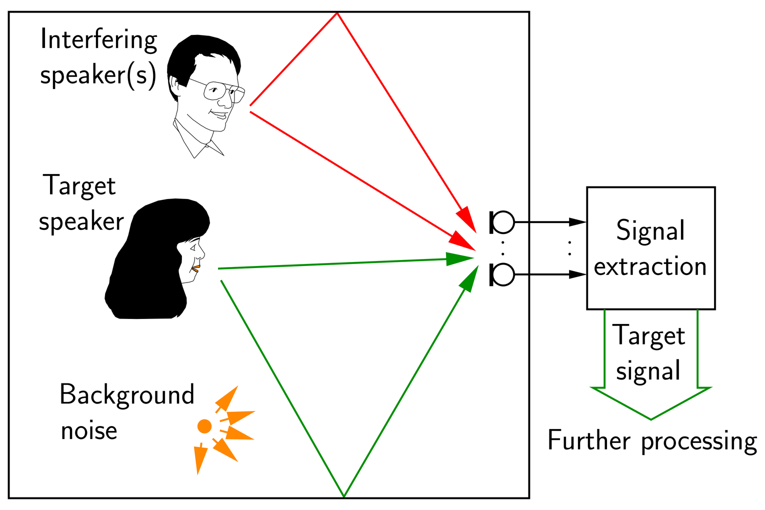
Improving Blind Source Separation Performance By Adaptive Array Geometries for Humanoid Robots

Hendrik Barfuss and *Walter Kellermann*

University of Erlangen-Nuremberg,
Multimedia Communications and Signal Processing,
Cauerstr. 7, D-91058 Erlangen, Germany
{barfuss,wk}@LNT.de

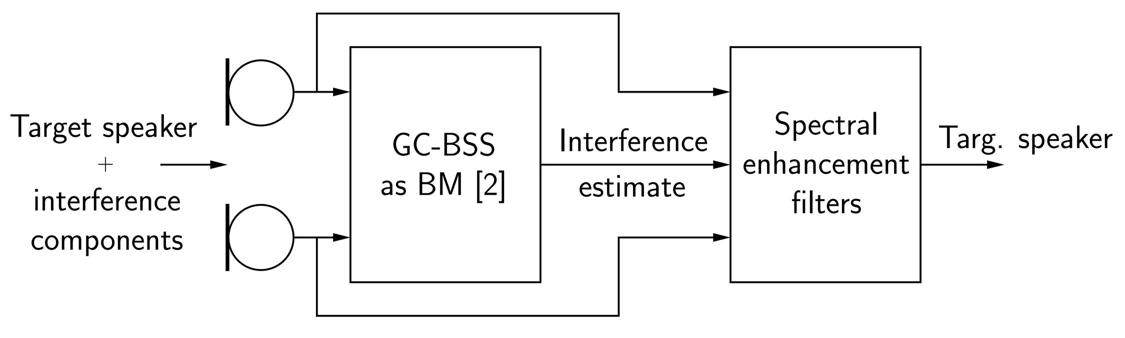
1. Introduction

Signal extraction scenario:



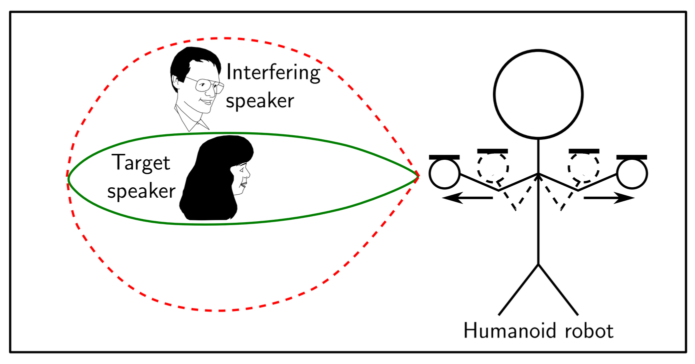
Goal: Extraction of desired source

Blind signal extraction (BSE) scheme [1]:



- Geometrically-constrained BSS (GC-BSS) acts as blocking matrix to estimate interference components
- Wiener-type spectral enhancement filters derived from interference estimate
- Extraction performance based on quality of interference estimate

Idea:



- Place microphones on movable limbs of a humanoid robot
- Use movable microphones to adapt microphone array topology to improve target signal suppression of GC-BSS
- Improved target signal suppression of GC-BSS leads to higher signal extraction performance of BSE [1]

2. Geometrically-constrained BSS

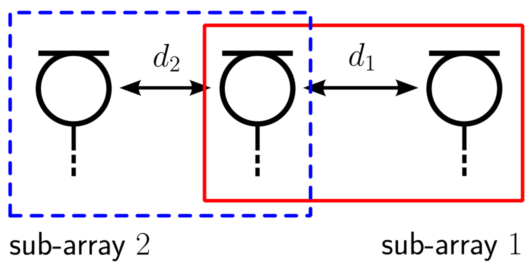
Cost function of GC-BSS [2]:

$$J_{\text{total}}(\mathbf{W}) = J_{\text{BSS}}(\mathbf{W}) + \gamma_{\text{C}} J_{\text{C}}(\mathbf{W}),$$

- $\mathbf{W}$: Matrix containing demixing filter weights
- $J_{\text{BSS}}(\mathbf{W})$: Generic cost function of TRINICON BSS [3]
- Minimization of $J_{\text{BSS}}(\mathbf{W})$ yields statistically independent output channels
- $J_{\text{C}}(\mathbf{W})$: Penalty term to force spatial null towards direction of target source in one output channel
- γ_{C}: Weight to control importance of penalty term, typically $0.4 < \gamma_{\text{C}} < 0.6$
- Direction of arrival ϕ_{t} of target source required for $J_{\text{C}}(\mathbf{W})$

3. Proposed Array Topology Adaptation Algorithm

Main concept:



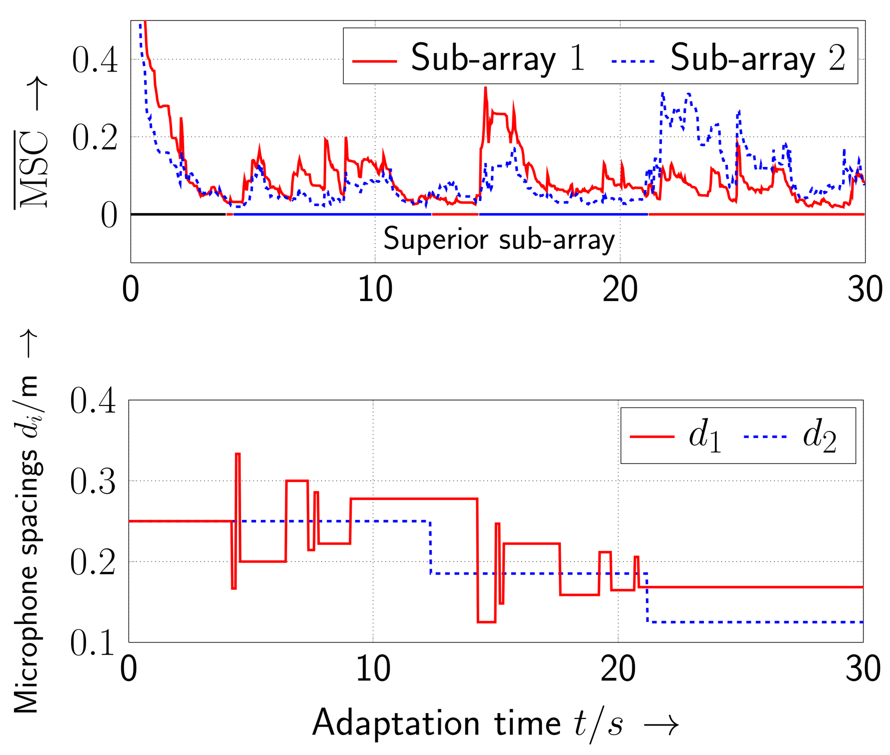
- Configuration of microphone array into two sub-arrays
- Use inferior sub-array to create new competitor for superior sub-array
- Adaptation of microphone spacing d_{inf} of inferior sub-array in dependence of microphone spacing d_{sup} of superior sub-array

$$d_{\text{inf}} = a_n d_{\text{sup}},\ n \in \{1, 2, 3, \ldots\},$$

with

$$a_n = \left(\frac{1+n}{2+n}\right)^{(-1)^{n+1}} = \left\{\frac{2}{3}, \frac{4}{3}, \frac{4}{5}, \frac{6}{5}, \frac{8}{7}, \ldots\right\}$$

Example of array topology adaptation:



- $\phi_{\text{t}} = 0°$, $\phi_{\text{int}} = 40°$, $T_{60} = 100\,\text{ms}$, Source distance: $1.0\,\text{m}$

4. Experiments

Evaluation measure:

- Weighted Magnitude Squared Coherence (MSC)

$$\overline{\text{MSC}} = \frac{1}{\sum_{\nu=0}^{\nu_{\max}} W(\nu)} \sum_{\nu=0}^{\nu_{\max}} W(\nu) \frac{|S_{y_1 y_2}(\nu)|^2}{S_{y_1 y_1}(\nu) \cdot S_{y_2 y_2}(\nu)},$$

with

$W(\nu)$: Window function

$S_{y_1 y_2}$, $S_{y_1 y_2}$, $S_{y_1 y_2}$: Auto- and cross-power spectrum densities of output channels

- Segmental target suppression gain

$$\Delta\overline{\text{TS}}_{\text{seg}} = \overline{\text{TS}}_{\text{seg},y_2} - \overline{\text{TS}}_{\text{seg},x}\ \text{dB},$$

with

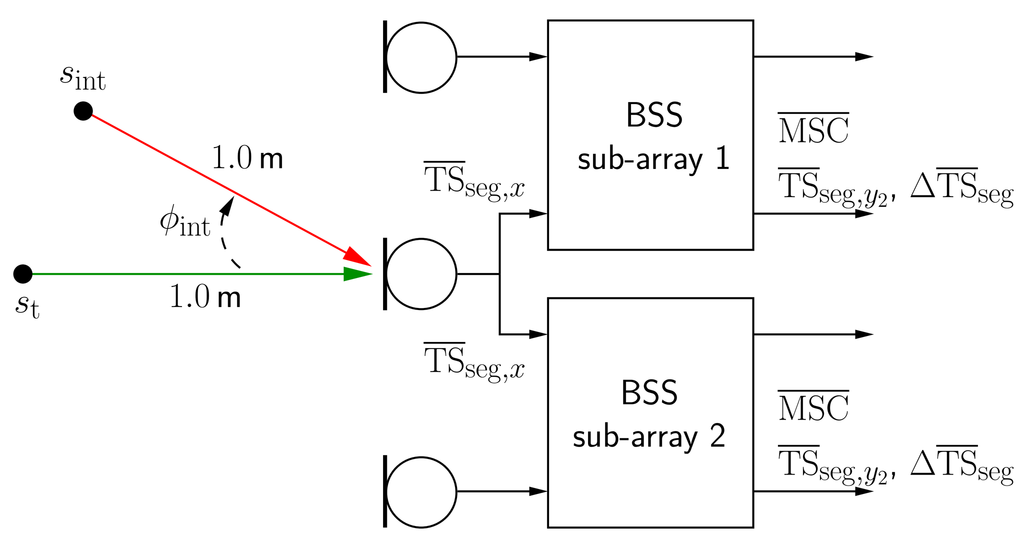
$$\overline{\text{TS}}_{\text{seg},y_2} = \frac{1}{K_{\text{S}}} \sum_{m=1}^{K_{\text{S}}} \left(10\log_{10}\left(\frac{\sum_{\kappa=1}^{N_{\text{S}}} y_{2,\text{int}}^2(\kappa + mN_{\text{S}})}{\sum_{\kappa=1}^{N_{\text{S}}} y_{2,\text{t}}^2(\kappa + mN_{\text{S}})}\right)\right)\ \text{dB}$$

$\overline{\text{TS}}_{\text{seg},x_2}$: Analogously to $\overline{\text{TS}}_{\text{seg},y_2}$

$y_{2,\text{t}}$, $y_{2,\text{int}}$: Target and interference components in second output channel

$N_{\text{S}} = 256, K_{\text{S}}$: Block length and number of blocks used to compute $\overline{\text{TS}}_{\text{seg},y}$ and $\overline{\text{TS}}_{\text{seg},y_2}$

Experimental Setup:



- $T_{60} = 100\,\text{ms}$, Source distance: $1.0\,\text{m}$
- $\phi_{\text{t}} = 0°$, $\phi_{\text{int}} \in \{20°, 40°, 60°, 80°\}$
- $f_s = 16\,\text{kHz}$, signal length: $60\,\text{s}$, Topology adaptation phase: $30\,\text{s}$
- Initial microphone spacings $d_i = 0.25\,\text{m}$
- Room impulse responses simulated using image method by Allen and Berkley [4]
- BSS: filter length $L = 512$, block length $N = 2L$

Results:

Adaptation	ϕ_{int}	$20°$	$40°$	$20°$	$80°$	Avg.
No	$\Delta\overline{\text{TS}}_{\text{seg}}$ [dB]	8.01	8.76	9.96	10.07	**9.20**
Yes	$\Delta\overline{\text{TS}}_{\text{seg}}$ [dB]	8.49	11.56	10.94	12.02	**10.75**

5. Conclusions

- Proposed array topology adaptation algorithm leads to improved separation, and thus, higher target source suppression performance
- By using array topology adaptation for humanoid robots, BSE algorithm [1] is expected to yield better signal extraction performance

References:

[1] K. Reindl, Y. Zheng, A. Schwarz, S. Meier, R. Maas, A. Sehr, and W. Kellermann, "A Stereophonic acoustic signal extraction scheme for noisy and reverberant environments," Computer Speech and Language (CSL), vol. 27, no. 3, pp. 726–745, May 2012.

[2] Y. Zheng, K. Reindl, and W. Kellermann, "BSS for improved interference estimation for blind speech signal extraction with two microphones," Int. Workshop on Comp. Advances in Multi-Sensor Adapt. (CAMSAP), Aruba, Dutch Antilles, pp. 253–256, Dec. 2009.

[3] H. Buchner, R. Aichner, and W. Kellermann, "TRINICON: A versatile framework for multichannel blind signal processing," IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP), vol. 3, pp. iii - 889-92, May 2004.

[4] J.B. Allen and D.A. Berkley, "Image method for efficiently simulating small-room acoustics," Journal Acoustic Society of America (JASA), vol. 65, no. 4, pp. 943–1086, April 1979.

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 609465.