AN ADAPTIVE MICROPHONE ARRAY TOPOLOGY FOR TARGET SIGNAL EXTRACTION WITH HUMANOID ROBOTS

Hendrik Barfuss and *Walter Kellermann*

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

ABSTRACT

In this paper, an unsupervised adaptation algorithm for the microphone array topology of a humanoid robot is proposed, so that the spatial filtering performance is improved. In the given exemplary case, the target suppression ('blocking') performance of a geometrically-constrained BSS (GC-BSS) algorithm is shown to improve by the adaptation of the array topology. As a decisive feature, an online performance measure for blind source separation is introduced which provides a robust and reliable estimate of the instantaneous signal separation performance based on currently observable data. Experimental results based on simulated environments confirm the efficacy of the concept.

Index Terms— Adaptive array topology, microphone arrays, humanoid robots, blind source separation, target suppression

1. INTRODUCTION

For natural human/robot interaction, robot audition should support speech communication even if the human is at a distance of several meters. As an example, we may consider a robot acting as an information point in public spaces, e.g., a welcoming robot in a hotel lobby. In such a scenario, the robot will be located at some distance from his desired human communication partner while other interfering speakers and background noise act as disturbance. Thus, a key problem in robot audition is to extract the desired source signal from the mixture of desired and interfering sources and background noise.

Using a humanoid robot offers the opportunity to place microphones not only on the head, but also on the movable limbs, allowing to change the aperture size of the microphone array, e.g., by letting the robot stretch out its arms or pull them back in. This movement could be incorporated into a welcoming gesture of the robot. This mechanism can be used to adapt the microphone spacings to the current acoustic scenario, and therefore, enable the robot to better focus on the desired source. An illustration of this concept is given in Fig. 1. The dashed red ellipsoid denotes the case where the robot cannot distinguish between the desired and interfering source due to an inappropriate array aperture. Adaptation of the microphone spacings enables the robot to focus on the desired source, as denoted by the solid green ellipsoid.

In [1], a two-channel Blind Signal Extraction (BSE) approach has been proposed, showing promising results in noisy living-room-like environments such as the one of the PASCAL CHiME challenge [2]. This BSE approach consists of two main blocks and is illustrated in Fig. 2. In the first block, interference components are estimated as described in [3]. The approach in [3] exploits GC-BSS as a Blocking Matrix (BM) for separating the desired (target) source signal components including correlated echoes from all undesired interference components. GC-BSS requires an estimate of the Direction of Arrival (DoA) of the desired signal which can be obtained as, e.g., described in [4, 5, 6]. In the second block of the BSE extraction scheme, Wiener-type spectral enhancement filters are derived from the interference estimate obtained from GC-BSS, and are applied to the microphone signals in order to suppress all undesired signal components in the output of the system. The key to obtain a good signal extraction performance is to obtain a good noise estimate, since the quality of the latter determines the performance of Wiener-type filters and, thus, of the entire extraction algorithm. As a decisive advantage of using GC-BSS for interference estimation, no source activity detection or estimation nor any source modeling is required, and no knowledge of the array topology is necessary. Only a rough estimate of the DoA of the target source is required.

In this paper, the concept of an algorithm is proposed which iteratively adapts the array topology of a linear microphone array, such that the target source suppression performance of GC-BSS is improved. As a key ingredient, an online performance measure is in-

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.

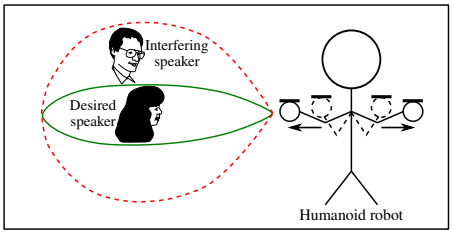


Fig. 1. Illustration of the underlying idea of the proposed adaptation algorithm.

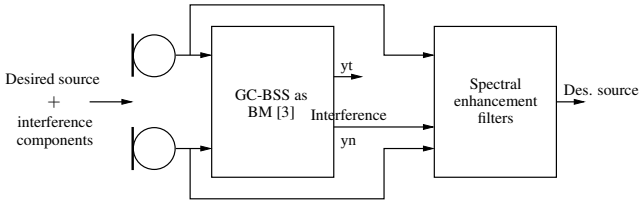


Fig. 2. Illustration of the BSE approach proposed in [1].

troduced, which estimates the target source suppression performance of GC-BSS blindly and reliably based on currently observable data. The efficacy of the proposed algorithm is verified by experiments using GC-BSS in simulated acoustic environments for a three-sensor linear microphone array.

The paper is structured as follows. Section 2 briefly outlines GC-BSS. In Section 3, the array topology adaptation algorithm is introduced. Employed performance measures, the experimental setup, and experimental results are presented in Section 4 and a summary and an outlook on future work is given in Section 5.

2. GEOMETRICALLY-CONSTRAINED BSS

The cost function of GC-BSS is given by [3]

$$J_{\text{total}}(\mathbf{W}) = J_{\text{BSS}}(\mathbf{W}) + \gamma_{\text{C}} J_{\text{C}}(\mathbf{W}), \tag{1}$$

where $J_{\text{BSS}}(\mathbf{W})$ is the generic cost function of the TRINICON (TRIple-N Independent COmponent Analysis for CONvolutive mixtures) Blind Source Separation (BSS) algorithm, introduced in [7]. $\mathbf{W}$ is the matrix containing all the filter weights of the set of demixing filters of the GC-BSS system. Minimizing $J_{\text{BSS}}(\mathbf{W})$ leads to a maximization of the statistical independence of the two output channels of the GC-BSS system. Since the BSE approach in [1] only requires two input channels, in this paper only GC-BSS with two input and two output channels is considered. In the remainder of this paper, the first and second output channel of the GC-BSS system are referred to as target output channel y_{t} and noise output channel y_{n}, respectively. The cost function of generic BSS is combined with an additional penalty term $J_{\text{C}}(\mathbf{W})$ weighted with a scalar γ_{C}, which can be chosen to control the importance of the geometric constraint, and is typically in the range of $0.4 < \gamma_{\text{C}} < 0.6$. $J_{\text{C}}(\mathbf{W})$ in (1) is chosen such that a spatial null is forced towards the DoA ϕ_{des} of the target source in the noise (second) output channel of GC-BSS. Since only the target source needs to be suppressed, only the noise output channel of GC-BSS is controlled by the penalty term. Please note that for the geometric constraint, we assume the DoA of the target source is known in this paper. The cost function $J_{\text{total}}(\mathbf{W})$ is formulated in the time domain, avoiding the problem of permutation ambiguities in different frequency bands. For the experiments, a second-order statistics realization of the GC-BSS algorithm has been used, see [3].

3. PROPOSED ARRAY TOPOLOGY ADAPTATION ALGORITHM

The main concept of the array topology adaptation algorithm is based on the fact that the microphone array can be configured as two sub-arrays. A good separation performance of the GC-BSS algorithm corresponds to well separated desired and interfering signal components. Thus, the adaptation of the two microphone spacings d_i, $i \in \{1, 2\}$ is based on the separation performance obtained after performing GC-BSS for each of the two sub-arrays. In the following, the array topology adaptation algorithm is presented by using a three-sensor linear array which is configured as two two-sensor sub-arrays, as illustrated in Fig. 3. For this presentation, the acoustic scenario is assumed to remain time-invariant. Suggested microphone positions are given by one microphone attached to each of the robot's hands and one microphone located at the robot's torso. Given this configuration, each sub-array consists of the center microphone which stays at a fixed position and a microphone mounted

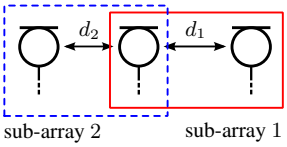


Fig. 3. Illustration of the employed sub-arrays for the case of a three-sensor linear array.

to the left hand or right hand, respectively. The two sub-arrays are illustrated by the solid red and dashed blue box in Fig. 3.

Without loss of generality, we assume that the initial microphone spacings are chosen such that $d_1 = d_2$. At the beginning of the array topology adaptation, an initial adaptation phase is required in order to let the two GC-BSS systems adapt to the current scenario. To this end, the input signal pairs are processed block-wise for each GC-BSS block m until convergence of the GC-BSS algorithm, see, e.g., the block-online adaptation of TRINICON [8]. A performance measure $f(d_i, m)$, $i \in \{1, 2\}$ is computed which characterizes the separation performance of the corresponding i-th sub-array with microphone spacing d_i after the m-th data block. It estimates the coherence between the two output channels of the GC-BSS system. Small values of $f(d_i, m)$ correspond to a good separation performance, and thus, to a good suppression of the target source in the noise output channel. The chosen performance measure is discussed in Section 4 in more detail. Convergence of the two GC-BSS systems is assumed when both systems yield a performance measure which is lower than a predefined threshold η for $m_{\eta,\max}$ consecutive GC-BSS data blocks, i.e., $f(d_i, m) < \eta$, $m \in \{1, ..., m_{\eta,\max}\}$, $i \in \{1, 2\}$.

After this initial adaptation phase, the currently inferior and superior sub-arrays are determined based on the obtained performance measure: The superior sub-array is chosen to be the sub-array yielding the lower performance measure and vice versa. After this, the microphone spacings of the two sub-arrays are adapted as follows: Based on the obtained performance measures $f(d_i, m)$ of the GC-BSS algorithm of each sub-array, the microphone spacing d_{inf} of the currently inferior sub-array is adapted depending on the microphone spacing d_{sup} of the currently superior sub-array:

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

where a_n is a scalar value taken from the sequence

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

The idea is to use the inferior sub-array to create a new competitor for the currently superior sub-array with a similar microphone spacing as the superior sub-array. Whenever a sub-array becomes the inferior sub-array, its microphone spacing is adapted according to (2) and (3), beginning with $n = 1$. If, after a new GC-BSS adaptation phase, this sub-array is still inferior, the d_{inf} is adapted with $n = n + 1$. In the case where the currently inferior sub-array becomes the superior sub-array, n is set to 1 and the spacing of the now inferior sub-array is adapted according to (2) and (3), starting at $n = 1$ again. The currently inferior sub-array becomes the new superior sub-array, if it yields a lower $f(d_i, m)$ than the superior sub-array for $m_{\alpha,\max}$ consecutive GC-BSS data blocks, i.e. if $f(d_{\text{inf}}, m) < \alpha f(d_{\text{sup}}, m)$, $m \in \{1, ..., m_{\alpha,\max}\}$, where $f(d_{\text{sup}}, m)$ and $f(d_{\text{inf}}, m)$ denote the performance measure obtained from the superior and inferior sub-array, respectively, after the m-th GC-BSS data block, and α is a scalar value in the range $0 < \alpha \leq 1$.

4. EXPERIMENTS

In this section, results from first experiments are presented. The chosen performance measures for GC-BSS are introduced in Subsection 4.1. The experimental setup is given in Subsection 4.2 and results are presented in Subsection 4.3.

4.1. Performance measures for GC-BSS

A very crucial aspect of the proposed adaptation algorithm is the need of a performance measure which is necessary to compare the separation performance of the two sub-arrays blindly and reliably. The used performance measure is based on the weighted Magnitude Squared Coherence (MSC) of the two output signals y_t and y_n of GC-BSS:

$$\overline{\mathrm{MSC}} = \frac{1}{\sum_{\nu=0}^{\nu_\mathrm{max}} W(\nu)} \sum_{\nu=0}^{\nu_\mathrm{max}} W(\nu) \frac{|S_{y_\mathrm{t} y_\mathrm{n}}(\nu)|^2}{S_{y_\mathrm{t} y_\mathrm{t}}(\nu) \cdot S_{y_\mathrm{n} y_\mathrm{n}}(\nu)}, \quad (4)$$

where ν_max denotes the maximum number of frequency bins ν, and $S_{y_\mathrm{t} y_\mathrm{n}}(\nu)$ and $S_{y_\mathrm{t} y_\mathrm{t}}(\nu)$ and $S_{y_\mathrm{n} y_\mathrm{n}}(\nu)$, represent the auto- and cross-PSDs of the target and noise output channel of the GC-BSS system, respectively. The weighting function $W(\nu)$ at each frequency bin ν of the MSC and is defined as

$$W(\nu) = \frac{S_{y_\mathrm{t} y_\mathrm{t}}(\nu) + S_{y_\mathrm{n} y_\mathrm{n}}(\nu)}{2}. \quad (5)$$

In Fig. 4, an exemplary weighting function is illustrated using a logarithmic y-scale. For the sake of clarity, instead of the frequency bins ν, the corresponding absolute frequency values are given in Fig. 4. The weighting function emphasizes low frequencies which contain more relevant speech information (pitch and formants) and are harder to separate. The $\overline{\mathrm{MSC}}$ is limited to the range of $0 \leq \overline{\mathrm{MSC}} \leq 1$, where $\overline{\mathrm{MSC}} = 0$ corresponds to statistically orthogonal signals.

In addition to the weighted MSC (4), the segmental target suppression gain (see [9]) for the noise output channel of each GC-BSS system is measured. It is defined as

$$\Delta\overline{\mathrm{TS}}_\mathrm{seg} = \overline{\mathrm{TS}}_{\mathrm{seg},y_\mathrm{n}} - \overline{\mathrm{TS}}_{\mathrm{seg},x}\,\mathrm{dB}, \quad (6)$$

where the segmental target suppression $\overline{\mathrm{TS}}_{\mathrm{seg},y_\mathrm{n}}$ in the noise output channel y_n of the GC-BSS system is defined as [9]

$$\overline{\mathrm{TS}}_{\mathrm{seg},y_\mathrm{n}} = \frac{1}{K_\mathrm{S}} \sum_{m=1}^{K_\mathrm{S}} \left(10\log_{10} \left(\frac{\sum_{\kappa=1}^{N_\mathrm{S}} y_{\mathrm{n,int}}^2(\kappa + mN_\mathrm{S})}{\sum_{\kappa=1}^{N_\mathrm{S}} y_{\mathrm{n,des}}^2(\kappa + mN_\mathrm{S})} \right) \right) \mathrm{dB}, \quad (7)$$

with $y_{\mathrm{n,des}}$ and $y_{\mathrm{n,int}}$ being the desired (target) and interference components in the noise output channel of each GC-BSS system, respectively. $\overline{\mathrm{TS}}_{\mathrm{seg},x}$ in (6) denotes the segmental target suppression at the center microphone of the employed three-element microphone array and is defined analogously to (7). The parameters $N_\mathrm{S} = 256$ samples and K_S in (7) denote the block length and the number of blocks used to compute the segmental target suppression. High values of $\Delta\overline{\mathrm{TS}}_\mathrm{seg}$ indicate a good suppression of the target source in the noise output channel. For the sake of simplicity, Fig. 5 includes an overview where $\overline{\mathrm{MSC}}$, $\Delta\overline{\mathrm{TS}}_\mathrm{seg}$, $\overline{\mathrm{TS}}_{\mathrm{seg},y_\mathrm{n}}$, and $\overline{\mathrm{TS}}_{\mathrm{seg},x}$ are computed in the system.

4.2. Experimental setup

The adaptation algorithm has been tested in a two-source environment with a male and a female speaker of equal power. The male speaker, who was the desired source, was always located at $\phi_\mathrm{des} = 0^\circ$, whereas the interfering speaker was located at different positions $\phi_\mathrm{int} \in \{20^\circ, 40^\circ, 60^\circ, 80^\circ\}$. Both sources were located at a distance of 1.0 m with respect to the center microphone of the microphone array, as illustrated in Fig. 5 for $\phi_\mathrm{int} = 40^\circ$. The Room Impulse Responses (RIRs), modeling the propagation from the sources to the microphones, were simulated using the image method proposed by Allen and Berkley [10]. The simulated room was of dimensions $(4.0\,\mathrm{m} \times 4.0\,\mathrm{m} \times 2.5\,\mathrm{m})$ with a reverberation time of $T_{60} \approx 100\,\mathrm{ms}$ and $T_{60} \approx 400\,\mathrm{ms}$, corresponding to a critical distance [11] of approximately 1.13 m and 0.56 m, respectively. The microphone signals were synthesized by convolving clean speech signals of sampling rate $f_\mathrm{s} = 16\,\mathrm{kHz}$ with the simulated RIRs. For the GC-BSS algorithms a filter length of $L = 512$ and $\gamma_\mathrm{C} = 0.5$ were used. The power-spectral densities, required for the calculation of $\overline{\mathrm{MSC}}$ (4), were estimated using the Welch method with a window length of $2L$ samples and 50% overlap. Across the blocks, a recursive averaging with a forgetting factor of $\lambda = 0.95$ has been applied for the estimated power-spectral densities. For array topology adaptation, the parameters were chosen as follows: $\eta = 0.05$, $\alpha = 0.95$, and $m_{\eta,\mathrm{max}} = m_{\alpha,\mathrm{max}} = 5$.

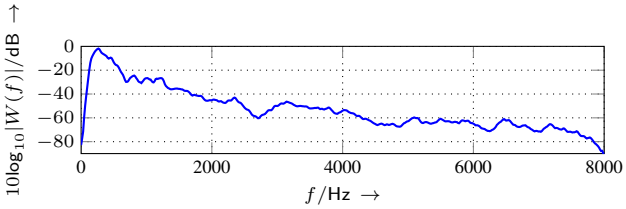


Fig. 4. Example of an average speech PSD used as window function (5) in (4).

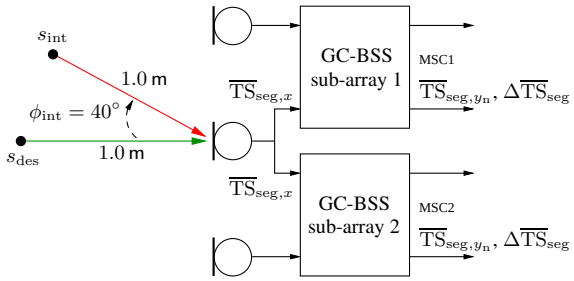
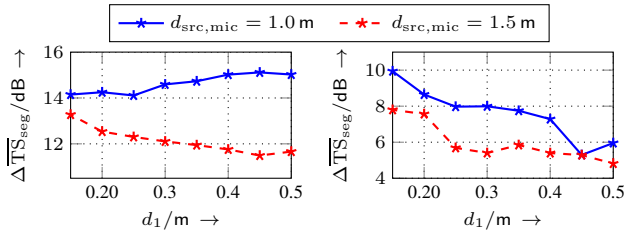


Fig. 5. Illustration of the multi-speaker scenario and employed measures of performance.

4.3. Experimental results

In Fig. 6, the obtained target suppression performance $\Delta\overline{\mathrm{TS}}_\mathrm{seg}$ in dB of the employed GC-BSS two-channel system for different microphone spacings d_1 from 0.15 m to 0.5 m and two different source-microphone distances $d_\mathrm{src,mic} \in \{1.0, 1.5\}$ m (measured relative to the center microphone), denoted by the solid blue and dashed red lines, for $T_{60} \approx 100\,\mathrm{ms}$ (Fig. 6(a)) and $T_{60} \approx 400\,\mathrm{ms}$ (Fig. 6(b)) is illustrated. The results were obtained from sub-array 1. The interfering source was located at $\phi_\mathrm{int} = 60^\circ$. It can be seen that for $T_{60} \approx 100\,\mathrm{ms}$ and $d_\mathrm{src,mic} = 1.0\,\mathrm{m}$, larger microphone spacings yield a better suppression of the desired source, whereas for $T_{60} \approx 100\,\mathrm{ms}$ with $d_\mathrm{src,mic} = 1.5\,\mathrm{m}$ and for $T_{60} \approx 400\,\mathrm{ms}$, smaller microphone spacings lead to a better suppression. Thus, an array

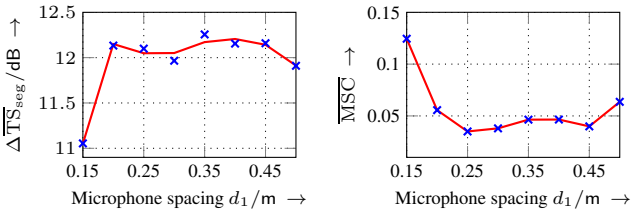


(a) $T_{60} \approx 100\,\mathrm{ms}$ (b) $T_{60} \approx 400\,\mathrm{ms}$

Fig. 6. Comparison of obtained target suppression performance $\Delta\overline{\mathrm{TS}}_{\mathrm{seg}}$ in dB of the employed GC-BSS two-channel system for varying microphone spacings d_1 and different source-microphone distances $d_{\mathrm{src,mic}}$ in two different acoustic environments.

topology adaptation mechanism seems desirable in order to adapt the microphone spacings to the acoustic scenario to improve suppression of the desired source, and, as a consequence, the overall performance of the BSE approach [1] described above.

In Fig. 7, a comparison between $\Delta\overline{\mathrm{TS}}_{\mathrm{seg}}$ and $\overline{\mathrm{MSC}}$ obtained from sub-array 1 for different microphone spacings d_1 between 0.15 and 0.5 m with $\phi_{\mathrm{int}} = 20^\circ$ and $T_{60} = 100\,\mathrm{ms}$ is given. The red curve is a polynomial of 6-th order that fits the data in the least squares sense. As can be seen, higher $\Delta\overline{\mathrm{TS}}_{\mathrm{seg}}$ values correspond to lower $\overline{\mathrm{MSC}}$ values and vice versa. Thus, the $\overline{\mathrm{MSC}}$ appears to be well suited to evaluate the separation performance of the GC-BSS algorithm.



(a) $\Delta\overline{\mathrm{TS}}_{\mathrm{seg}}$ (b) $\overline{\mathrm{MSC}}$

Fig. 7. Comparison of $\mathrm{SIR}_{\mathrm{mean}}$ and $\overline{\mathrm{MSC}}$ obtained from sub-array 1 for different microphone spacings.

In the following the target suppression performance with and without array topology adaptation is discussed. The results were obtained by using continuous speech signals of length 60 seconds. In the case of an array topology adaptation, the adaptation is performed using the first 30 seconds of the speech signal and the target suppression performance is evaluated using the second 30 seconds of the speech signal. In the case of no array topology adaptation, GC-BSS is only applied to the second 30 seconds of the input signals in order to evaluate the target suppression performance. For adaptive and non-adaptive array topology, the initial microphone spacings of both sub-arrays were set to 0.25 m and lower and upper bound of the array aperture were set to 0.125 m and 0.5 m, respectively.

Fig. 8 includes an illustration of an exemplary array topology adaptation obtained with $\phi_{\mathrm{int}} = 40^\circ$ and $T_{60} \approx 100\,\mathrm{ms}$. In the upper plot, the $\overline{\mathrm{MSC}}$ of each sub-array is illustrated, where the colored bar at the bottom of the upper plot denotes the currently superior sub-array. The lower plot shows the development of the two microphone spacings over adaptation time.

In Table 1, a comparison of $\Delta\overline{\mathrm{TS}}_{\mathrm{seg}}$ for both non-adaptive and adaptive array topologies for a reverberation time of $T_{60} \approx 100\,\mathrm{ms}$

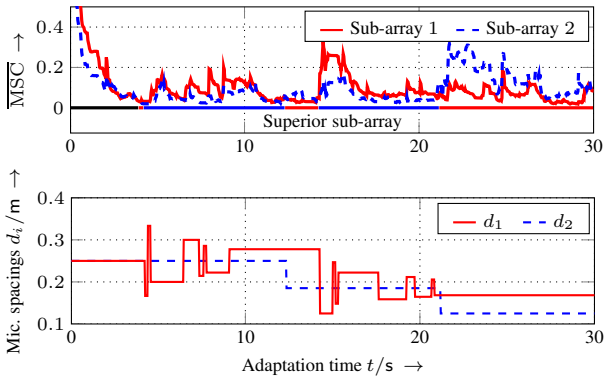


Fig. 8. Illustration of array topology adaptation for $\phi_{\mathrm{int}} = 40^\circ$ and $T_{60} \approx 100\,\mathrm{ms}$.

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

Table 1. Obtained $\Delta\overline{\mathrm{TS}}_{\mathrm{seg}}$ without and with topology adaptation, $T_{60} \approx 100\,\mathrm{ms}$.

is given. The $\Delta\overline{\mathrm{TS}}_{\mathrm{seg}}$ values presented here are the values of the better performing sub-array. As can be seen, array topology adaptation yields better target suppression performance than no topology adaptation for all simulated scenarios. By employing array topology adaptation, on average 1.55 dB of target suppression were gained.

For $T_{60} \approx 400\,\mathrm{ms}$, array topology adaptation yields better results for all tested scenarios, except $\phi_{\mathrm{int}} = 20^\circ$ where no better array topology could be found, as presented in Table 2. On average, target suppression was improved by 1.06 dB.

Adaptation	ϕ_{int}	20°	40°	20°	80°	Avg.
No	$\Delta\overline{\mathrm{TS}}_{\mathrm{seg}}$ [dB]	6.74	6.33	6.98	7.54	**6.90**
Yes	$\Delta\overline{\mathrm{TS}}_{\mathrm{seg}}$ [dB]	6.74	7.85	9.19	8.04	**7.96**

Table 2. Obtained $\Delta\overline{\mathrm{TS}}_{\mathrm{seg}}$ without and with topology adaptation, $T_{60} \approx 400\,\mathrm{ms}$.

5. SUMMARY AND OUTLOOK

In this paper, the generic concept of an adaptive microphone array topology was introduced for application with humanoid robots. It was demonstrated that this is an efficient method to improve the desired source suppression performance of GC-BSS. By using this adaptation algorithm in the context of the BSE scheme [1], it is expected to provide better noise estimates and thus increase the signal extraction performance. Consequently, Automatic Speech Recognition (ASR) scores are expected to improve significantly in complex and adverse acoustic scenarios, leading to an according reduction of task completion time in human/robot interaction. Future work will include a more detailed investigation and refinements of the proposed performance measure, robust optimization of the current adaptation mechanism with recorded microphone data in different acoustic environments, as well as the generalization to a larger number of movable microphones.

6. 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] J. Barker, E. Vincent, N. Ma, H. Christensen, and P. Green, "The PASCAL CHiME speech separation and recognition challenge.," *Computer Speech and Language (CSL)*, vol. 27, no. 3, pp. 621–633, May 2013.

[3] Y. Zheng, K. Reindl, and W. Kellermann, "BSS for improved interference estimation for blind speech signal extraction with two microphones," in *Proc. 3rd International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, December 2009, pp. 253–256.

[4] C. Knapp and G.C. Carter, "The generalized correlation method for estimation of time delay," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 24, no. 4, pp. 320–327, August 1976.

[5] F. Nesta, P. Svaizer, and M. Omologo, "Robust two-channel TDOA estimation for multiple speaker localization by using recursive ICA and a state coherence transform," in *Acoustics, Speech and Signal Processing (ICASSP). IEEE International Conference on*, April 2009, pp. 4597–4600.

[6] A. Lombard, Y. Zheng, H. Buchner, and W. Kellermann, "TDOA estimation for multiple sound sources in noisy and reverberant environments using broadband independent component analysis," *IEEE Transactions on Audio, Speech, and Language Processing (ASLP)*, vol. 19, no. 6, pp. 1490 – 1503, August 2011.

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

[8] R. Aichner, H. Buchner, F. Yan, and W. Kellermann, "A real-time blind source separation scheme and its application to reverberant and noisy acoustic environments," *Signal Processing*, vol. 86, pp. 1260–1277, June 2006.

[9] R. Aichner, *Acoustic Blind Source Separation in Reverberant and Noisy Environments, PhD thesis*, University Erlangen-Nuremberg, Germany, October 2007.

[10] 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, April 1979.

[11] J. Blauert and N. Xiang, *Acoustics for Engineers: Troy Lectures*, Springer, 2008.