# HRTF-BASED ROBUST LEAST-SQUARES FREQUENCY-INVARIANT BEAMFORMING

*Hendrik Barfuss, Christian Huemmer, Gleni Lamani, Andreas Schwarz, and Walter Kellermann*

Multimedia Communications and Signal Processing,
Friedrich-Alexander University Erlangen-Nürnberg
Cauerstr. 7, 91058 Erlangen, Germany
{barfuss,huemmer,schwarz,wk}@LNT.de, g.lamani@hotmail.com

## ABSTRACT

In this work, a Head-Related Transfer Function (HRTF)-based Robust Least-Squares Frequency-Invariant (RLSFI) beamformer design is proposed. The HRTF-based RLSFI beamformer accounts for the influence of a robot's head on the sound field. The performance of the new HRTF-based RLSFI beamformer is evaluated using signal-based measures and word error rates for an off-the-shelf speech recognizer, and is compared to the performance of the original free-field RLSFI beamformer design. The experimental results confirm the efficacy of the proposed HRTF-based beamformer design for robot audition.

***Index Terms***— Spatial filtering, robust superdirective beamforming, white noise gain, signal enhancement, robot audition

## 1. INTRODUCTION

A key problem in robot audition is to extract the desired source signal from a mixture of desired and interfering sources as well as background noise. For this purpose, beamforming techniques can be employed.

Typically, beamformers are designed assuming free-field propagation of sound waves. However, when applied to a microphone array mounted at a robot's head, these beamformers do not account for the influence of the robot's head on the sound field, e.g., attenuation of sound waves, which reduces the signal enhancement performance. To overcome this limitation, Head-Related Transfer Functions (HRTFs) can be incorporated into the beamformer design by substituting the free-field steering vectors with HRTFs.

In [1], this strategy has been used to improve the separation performance of a geometrically constrained source separation algorithm. In [2, 3, 4], a set of HRTF-based Minimum Variance Distortionless Response (MVDR) beamformers, designed under the assumption of a diffuse noise field, has been used as a pre-processing stage for a blind source separation algorithm. In all cases, the source separation performance was increased.

In this contribution, a Robust Least-Squares Frequency-Invariant (RLSFI) beamformer, originally proposed in [5], is extended by incorporating HRTFs instead of free-field steering vectors into the beamformer design. The HRTF-based RLSFI beamformer accounts for the influence of the robot's head onto the sound field and allows for a direct control of the tradeoff between the beamformer's directivity and robustness. Thus, it is very well suited for the application to robot audition. The proposed HRTF-based RLSFI beamformer is compared to the original free-field RLSFI beamformer by evaluating frequency-weighted segmental Signal-to-Noise Ratio (fwSegSNR) values and Word Error Rates (WERs) of an Automatic Speech Recognition (ASR) system in a multi-speaker scenario.

The remainder of this work is structured as follows: In Section 2.1, the original beamformer design is briefly summarized and in Section 2.2 the new HRTF-based design is introduced. Section 3 contains an evaluation and comparison of the signal enhancement performance of both beamformer designs. A conclusion and an outlook to future work is given in Section 4.

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 and from the Deutsche Forschungsgemeinschaft (DFG) under contract number KE 890/4-2.

## 2. ROBUST FILTER-AND-SUM BEAMFORMING

### 2.1. Robust least-squares frequency-invariant (RLSFI) beamforming

In Fig. 1, the block diagram of a Filter-and-Sum Beamformer (FSB) is illustrated. The output signal $y[k]$ at time instant $k$ is obtained by convolving the microphone signals $x_n[k]$, $n \in \{0, \dots, N-1\}$ with Finite Impulse Response (FIR) filters $\mathbf{w}_n = [w_{n,0}, \dots, w_{n,L-1}]^T$ of length $L$, followed by a summation over all $N$ channels. Column vector $\mathbf{p}_n$ represents the position of the $n$-th microphone in Cartesian coordinates. The beamformer response of an FSB as depicted in Fig. 1 is given as [5, 6, 7]:

$$B(\omega, \phi, \theta) = \sum_{n=0}^{N-1} W_n(\omega) g_n(\omega, \phi, \theta), \tag{1}$$

where $W_n(\omega) = \sum_{l=0}^{L-1} w_{n,l} e^{-j\omega l}$ is the Discrete-Time Fourier Transform (DTFT) representation of $\mathbf{w}_n$. The sensor response of the $n$-th sensor to a plane wave with frequency $\omega$ traveling in the direction $(\phi, \theta)$ is represented by $g_n(\omega, \phi, \theta)$, with $\phi$ and $\theta$ denoting azimuth and elevation angle, respectively, defined as in [7]. Assuming free-field propagation, the sensor response

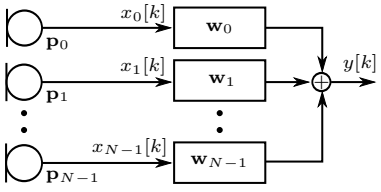



**Fig. 1**. Illustration of filter-and-sum beamforming [6].

$g_n(\omega,\phi,\theta) = e^{-j\mathbf{k}^T\mathbf{p}_n}$ is a free-field steering vector with wave vector $\mathbf{k} = -\frac{\omega}{c}[\sin(\theta)\cos(\phi), \sin(\theta)\sin(\phi), \cos(\theta)]^T$, and speed of sound $c$. Operator $(\cdot)^T$ denotes the transpose of a vector or matrix.

In [5], the design of a RLSFI FSB was proposed that approximates a desired beamformer response $\hat{B}(\omega,\phi,\theta)$ at each frequency $\omega$ in the Least-Squares (LS) sense subject to a distortionless response constraint in the desired look direction and a constraint on the White Noise Gain (WNG). In order to obtain a numerical solution, the LS approximation is performed for a discrete set of $P$ frequencies $\omega_p$, $p \in \{0,\ldots,P-1\}$ and $M$ look directions $(\phi_m,\theta_m)$, $m \in \{0,\ldots,M-1\}$. The optimization problem of the RLSFI FSB formulated in matrix notation is given as [5]

$$\underset{\mathbf{w}_\mathrm{f}(\omega_p)}{\operatorname{argmin}} \|\mathbf{G}(\omega_p)\mathbf{w}_\mathrm{f}(\omega_p) - \hat{\mathbf{b}}\|_2^2 \tag{2}$$

subject to constraints on the WNG and the response in desired look direction, respectively:

$$\frac{|\mathbf{w}_\mathrm{f}^T(\omega_p)\mathbf{d}(\omega_p)|^2}{\mathbf{w}_\mathrm{f}^H(\omega_p)\mathbf{w}_\mathrm{f}(\omega_p)} \geq \gamma > 0, \quad \mathbf{w}_\mathrm{f}^T(\omega_p)\mathbf{d}(\omega_p) = 1, \tag{3}$$

where $\hat{\mathbf{b}} = [\hat{B}(\phi_0,\theta_0),\ldots,\hat{B}(\phi_{M-1},\theta_{M-1})]^T$ is a vector containing the desired responses for all look directions, matrix $[\mathbf{G}(\omega_p)]_{mn}{=}e^{-j\mathbf{k}_m^T\mathbf{p}_n}$, $\mathbf{w}_\mathrm{f}(\omega_p){=}[W_0(\omega_p),\ldots,W_{N-1}(\omega_p)]^T$, and $\mathbf{d}(\omega_p) = [e^{-j\mathbf{k}_\mathrm{d}^T\mathbf{p}_0},\ldots,e^{-j\mathbf{k}_\mathrm{d}^T\mathbf{p}_{N-1}}]^T$ is the steering vector corresponding to the desired look direction $(\phi_\mathrm{d},\theta_\mathrm{d})$ with respective wave vector $\mathbf{k}_\mathrm{d}$. Operator $(\cdot)^H$ denotes transposition of conjugate complex vectors or matrices. Since the same desired response is chosen for all frequencies, the entries of $\hat{\mathbf{b}}$ are frequency-independent [5]. Equation (2) represents the LS approximation of the desired response. The first part of (3) represents the WNG constraint, where $\gamma$ is the lower bound on the WNG, which has to be defined by the user. The second part of (3) describes the distortionless response constraint which ensures that the desired signal coming from direction $(\phi_\mathrm{d},\theta_\mathrm{d})$ remains undistorted. The optimization problem in (2),(3) has to be solved for each frequency $\omega_p$ separately.

Fig. 2 illustrates an example of the RLSFI beamformer design described above. The beamformer was designed for $N = 5$ microphones mounted at the head of a humanoid robot as illustrated in Fig 6(b). The design was carried out for a frequency range $300\,\mathrm{Hz} \leq f \leq 5000\,\mathrm{Hz}$ (chosen with the application to speech signal capture in mind) and for two different WNG constraint values $\gamma_\mathrm{dB} = 10\log_{10}(\gamma) \in \{-25,-15\}\,\mathrm{dB}$. For the FIR approximation, a filter length $L = 1024$ was used. For the free-field steering vectors, the position of the microphones in Cartesian coordinates is required, which is known due to the fixed array geometry. Figs. 2a) and 2b) show the resulting beampatterns $\mathrm{BP}_\mathrm{dB}(\ldots) = 10\log_{10}(|B(\ldots)|^2)$ and Fig. 2c) illustrates the corresponding WNG. The beampatterns were evaluated with $g_n(\omega,\phi,\theta) = e^{-j\mathbf{k}^T\mathbf{p}_n}$, i.e., by using the free-field steering vectors as sensor responses. The main beam was formed towards broadside ($\phi_\mathrm{d} = 90°$) at an elevation of $\theta_\mathrm{d} = 56°$, with the elevation accounting for the small size of the robot relative to that of a human communication partner. The two beampatterns show a good spatial selectivity across the desired frequency range. However, the main beam broadens for lower frequencies. A higher WNG constraint $\gamma_\mathrm{dB}$ leads to a broader beam at lower frequencies, similar to a delay-and-sum beamformer with uniform filter weights, which is the most robust beamformer [5, 6]. Thus, the WNG constraint gives the user direct control over the tradeoff between directivity and robustness of the beamformer. Both designs fulfill the required WNG with some slight deviations, which are due to the FIR approximation of the optimum filter coefficients.

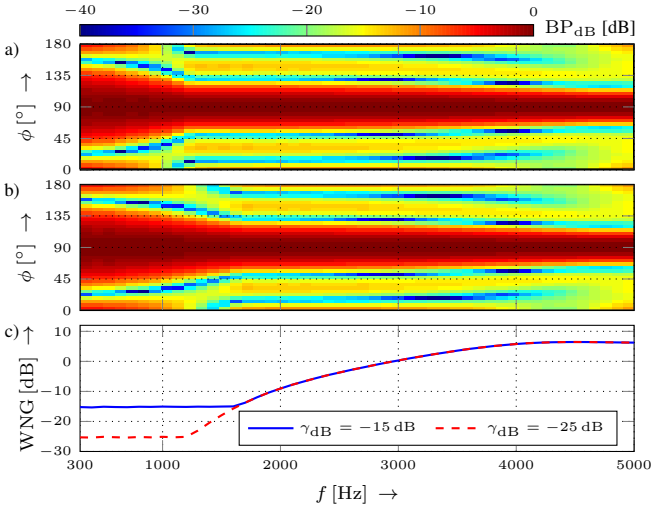



**Fig. 2**. Design example of free-field-based RLSFI beamformer for a five-element microphone array. Beampatterns for WNG constraints a) $\gamma_\mathrm{dB} = -25\,\mathrm{dB}$ and b) $\gamma_\mathrm{dB} = -15\,\mathrm{dB}$. The resulting WNG in illustrated in c).

### 2.2. HRTF-based robust least-squares frequency-invariant (RLSFI) beamforming

The steering vectors in $\mathbf{G}(\omega_p)$ and $\mathbf{d}(\omega_p)$ in (2) and (3), respectively, rely on the assumption of free-field propagation of sound waves. Thus, the beamformer design presented in Section 2.1 does not take the effect of the robot's head on the sound field into account. The consequence is illustrated in Fig. 3 which shows the beampattern of the free-field-based RLSFI beamformer design with $\gamma_\mathrm{dB} = -15\,\mathrm{dB}$, calculated with measured sensor responses $g_n(\omega,\phi,\theta) = h_{mn}(\omega)$, where $h_{mn}(\omega)$ is the measured HRTF between the $m$-th source position and $n$-th sensor at frequency $\omega$.
The spatial selectivity is drastically reduced as can be seen from comparison of Fig. 3 with Fig. 2b). In addition, no distortionless response in the desired look direction is maintained anymore.

To overcome this limitation, an HRTF-based RLSFI beamformer design is presented in the following, where the free-field steering vectors in the optimization problem (2),(3) are substituted with a set of HRTFs. The HRTF-based RLSFI optimization problem can then be formulated as

$$\underset{\mathbf{w}_\mathrm{f}(\omega_p)}{\operatorname{argmin}} \|\mathbf{G}_\mathrm{HRTF}(\omega_p)\mathbf{w}_\mathrm{f}(\omega_p) - \hat{\mathbf{b}}\|_2^2 \tag{4}$$

subject to

$$\frac{|\mathbf{w}_\mathrm{f}^T(\omega_p)\mathbf{d}_\mathrm{HRTF}(\omega_p)|^2}{\mathbf{w}_\mathrm{f}^H(\omega_p)\mathbf{w}_\mathrm{f}(\omega_p)} \geq \gamma > 0, \quad \mathbf{w}_\mathrm{f}^T(\omega_p)\mathbf{d}_\mathrm{HRTF}(\omega_p) = 1, \tag{5}$$

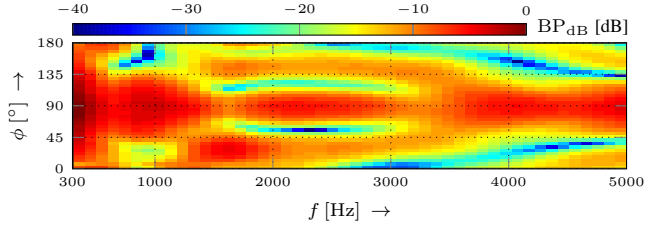



**Fig. 3**. Illustration of the beampattern in dB of the five-element free-field-based RLSFI beamformer with WNG constraint $\gamma_\mathrm{dB} = -15\,\mathrm{dB}$, computed with HRTFs as sensor responses $g_n(\omega,\phi,\theta)$.

where

$$[\mathbf{G}_{\mathrm{HRTF}}(\omega_p)]_{mn} = h_{mn}(\omega_p) \tag{6}$$

includes all HRTFs $h_{mn}(\omega_p)$, and

$$\mathbf{d}_{\mathrm{HRTF}}(\omega_p) = \left[h_{\mathrm{d}0}(\omega_p), \quad h_{\mathrm{d}1}(\omega_p), \quad \ldots, \quad h_{\mathrm{d}N-1}(\omega_p)\right]^T \tag{7}$$

includes the HRTFs between the source located at the desired look direction $(\phi_\mathrm{d}, \theta_\mathrm{d})$ and all $N$ microphones at frequency $\omega_p$. The HRTF-based design no longer requires knowledge of the microphone positions, but the HRTFs in $\mathbf{G}_{\mathrm{HRTF}}(\omega_p)$ and $\mathbf{d}_{\mathrm{HRTF}}(\omega_p)$ need to be measured before the beamformer is designed. The optimization problem in (4) and (5) is also convex and is solved using CVX.

A design example of the HRTF-based RLSFI beamformer is shown in Fig. 4. The same parameters as for the design example in Fig. 2 were used. The beamformer was steered towards $(\phi_\mathrm{d}, \theta_\mathrm{d}) = (90°, 56°)$. Note that as opposed to the examples in Figs. 2 and 3 the illustrated beamformer design is now based on HRTFs, which were measured for all $M$ look directions using maximum length sequences [8]. For calculating beampatterns and WNG, the measured HRTFs were used as sensor responses, as for the beampattern presented in Fig. 3. Similar to the free-field-based design, the two beampatterns show a good spatial selectivity across the desired frequency range. A comparison of Figs. 3 and 4b) shows that the HRTF-based RLSFI beamformer now produces a good spatial selectivity under realistic conditions. By comparing the design examples in Figs. 2 and 4, it can be seen that the HRTF-based design yields a broader main beam at lower frequencies for $\gamma_\mathrm{dB} = -15\,\mathrm{dB}$ than the free-field design. Also, the WNG of the HRTF-based design is lower than for the free-field design at higher frequencies. This confirms that the HRTF-based design cannot fully compensate the influence of the robot's head.

A closer look reveals that using HRTFs instead of free-field steering vectors for the beamformer design has the effect of limiting the maximum WNG that can be achieved by the resulting beamformer weights. The limitation is due to the fact that HRTFs usually have lower amplitudes than free-field steering vectors. Thus, in order to maintain the distortionless response of the beamformer in the

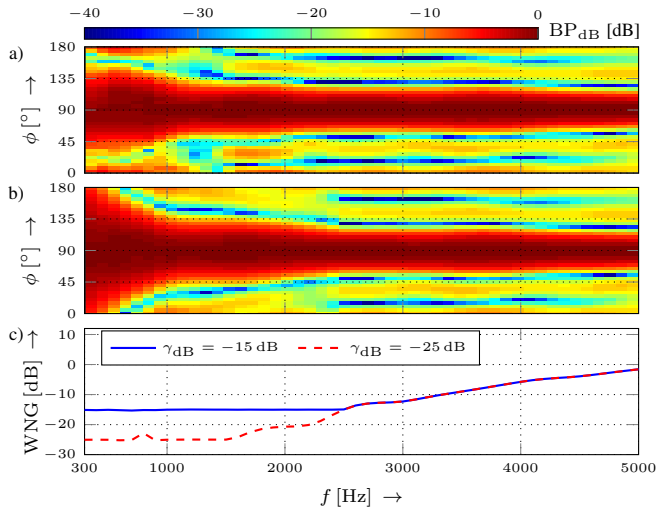



**Fig. 4**. Design example of HRTF-based RLSFI beamformer for a five-element microphone array. Beampatterns for WNG constraints a) $\gamma_\mathrm{dB} = -25\,\mathrm{dB}$ and b) $\gamma_\mathrm{dB} = -15\,\mathrm{dB}$. The resulting WNG is illustrated in c).

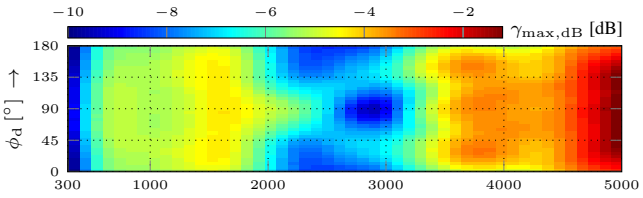



**Fig. 5**. Illustration of maximum possible WNG $\gamma_{\max}(\omega_p)$ for the HRTF-based beamformer design. In general, a lower maximum WNG is obtained than for the free-field design.

desired look direction, larger filter weights are needed to compensate for the lower amplitude of the HRTFs, which then results in a lower WNG. The maximum possible WNG $\gamma_{\max}$ of the HRTF-based RLSFI beamformer can be determined by solving the following optimization problem:

$$\gamma_{\max}(\omega_p) = \max_{\mathbf{w}_\mathrm{f}(\omega_p)} \frac{|\mathbf{w}_\mathrm{f}^T(\omega_p)\mathbf{d}_{\mathrm{HRTF}}(\omega_p)|^2}{\mathbf{w}_\mathrm{f}^H(\omega_p)\mathbf{w}_\mathrm{f}(\omega_p)} \tag{8}$$

subject to

$$\mathbf{w}_\mathrm{f}^T(\omega_p)\mathbf{d}_{\mathrm{HRTF}}(\omega_p) = 1, \tag{9}$$

which is also a convex optimization problem. For free-field steering vectors, the outcome of this optimization problem is always $\gamma_{\max}(\omega_p) = N$. In Fig. 5, $\gamma_{\mathrm{max,dB}}(\omega_p)$ is illustrated for a frequency range of $300\,\mathrm{Hz} \leq f \leq 5000\,\mathrm{Hz}$ and for desired look directions $0° \leq \phi_\mathrm{d} \leq 180°$ at an elevation angle of $\theta_\mathrm{d} = 56°$. The microphone array configuration was the same as for the previous design examples illustrated in Figs. 2 - 4. It can be seen that the maximum possible WNG $\gamma_{\mathrm{max,dB}}$ of the HRTF-based RLSFI beamformer is lower than $10\log_{10}(N = 5) \approx 7\,\mathrm{dB}$ of the free-field-based design for all directions and frequencies. Moreover, it is both look-direction- and frequency-dependent. A strategy to overcome the limitation of the WNG is to remove the distortionless response constraint from the optimization problem in (4) and (5). The spatial directivity would still be preserved to some extent due to the desired response that is approximated in the LS sense. This strategy, however, is not analyzed within this work but will be part of future studies.

## 3. SIGNAL ENHANCEMENT EVALUATION

In the following, the free-field- and HRTF-based RLSFI beamformer are compared with respect to their signal enhancement performance. At first, the experimental setup is introduced, followed by the presentation of the results of the signal enhancement evaluation.

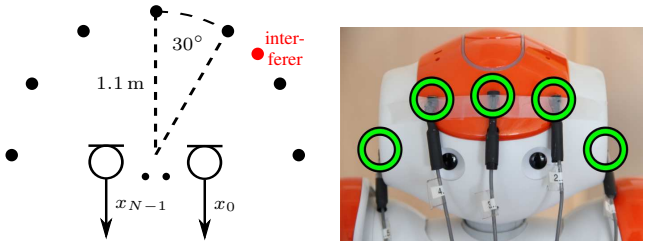
### 3.1. Experimental setup

The overall quality of the enhanced signal at the beamformer output was evaluated using Word Error Rates (WERs) of an automatic speech recognizer. The ASR engine PocketSphinx [9] was used with an acoustic model trained on clean speech from the GRID corpus [10], using MFCC+$\Delta$+$\Delta\Delta$ features and cepstral mean normalization. For the computation of the WER scores, only the letter and the number in the utterance were evaluated, as in the CHiME challenge [11]. The test signal contained 200 utterances. Moreover, the fwSegSNR as defined in [12] was evaluated. For calculating the input and output fwSegSNR, the desired signal at the center microphone and at the beamformer output were used as reference signal, respectively.

The two beamformer designs were evaluated in a two-speaker scenario where one interfering speech source was always located at

$\phi_{\mathrm{int}} = 45°$, and the desired speech source was located at positions between $\phi_{\mathrm{d}} = 0°$ and $\phi_{\mathrm{d}} = 180°$ in steps of $30°$, as illustrated in Fig. 6(a). All sources were located at an elevation of $\theta_{\mathrm{d}} = 56°$. For all experiments, a Signal-to-Interference Ratio (SIR) of $0\,\mathrm{dB}$ was considered, and no additional noise was added to the microphone signals. Moreover, the exact Direction-of-Arrival (DoA) of the desired source was assumed to be known.

The microphone signals were generated by convolving clean-speech sources with Room Impulse Responses (RIRs), which were measured in a lab room with a reverberation time of $T_{60} = 190\,\mathrm{ms}$ and a critical distance [13] of approximately $1.2\,\mathrm{m}$, using maximum length sequences. The source-robot distance was $1.1\,\mathrm{m}$, and the relative height of the source with respect to the robot's head was $0.73\,\mathrm{m}$, simulating a taller human which is interacting with the NAO robot which is of height $0.57\,\mathrm{m}$ [14]. Fig. 6(b) illustrates the employed microphone positions at the robot's head. The set of prototype HRTFs



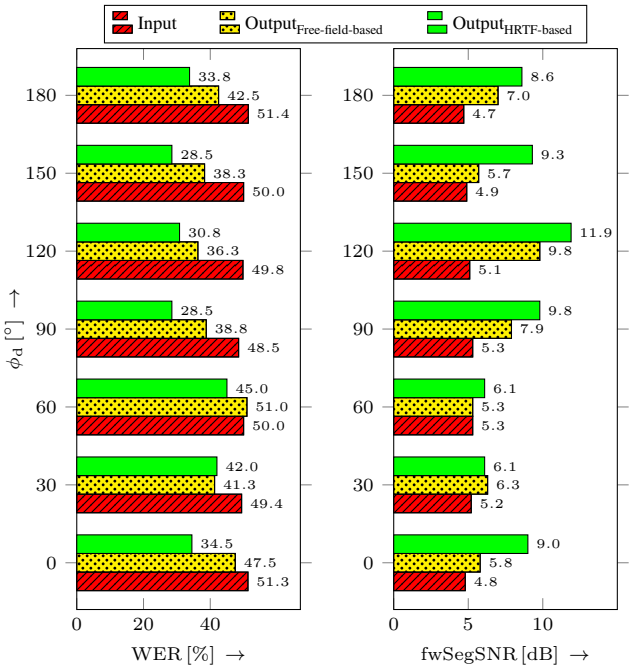

(a) Source positions. (b) Microphone positions.

**Fig. 6**. Illustration of evaluated source positions and the employed microphone positions at the robot's head.

which is required for the HRTF-based beamformer design was measured for the same microphone configuration and source positions as for the RIR measurements. The measurements were carried out for the robot looking towards broadside ($\phi = 90°$), at an elevation of $\theta = 90°$.

### 3.2. Results

In Figs. 7(a) and 7(b), the WERs and the fwSegSNR, obtained at the input and at the output of the free-field- and HRTF-based beamformers are illustrated. The results show that for $\phi_{\mathrm{d}} \in \{30°, 60°\}$, both beamformers have a limited performance, which is due to the close interfering source at $\phi_{\mathrm{int}} = 45°$. For most of the scenarios, the WER at the output is significantly decreased by both beamformer designs, with WER reductions of up to 21.5 and 13.5 percentage points for the HRTF-based and free-field-based beamformer, respectively. Most importantly, the HRTF-based beamformer yields significantly lower WERs in almost all scenarios than the free-field-based design, with relative WER reductions between the free-field- and HRTF-based beamformer of up to 13 percentage points. The obtained fwSegSNR values in Fig 7(b) show a similar behaviour. Both beamformers significantly improve the input signal by $6.8$ and $4.7\,\mathrm{dB}$ at the most, for the HRTF-based and free-field-based beamformer, respectively. For $\phi_{\mathrm{d}} \in \{30°, 60°\}$, the signal enhancement performance of both beamformers is similar. Analogously to the WERs, the HRTF-based design clearly outperforms the free-field-based design in the remaining scenarios, with relative fwSegSNR improvements of up to $3.6\,\mathrm{dB}$ between the free-field- and HRTF-based design, respectively.




(a) Obtained WER in %. (b) Obtained fwSegSNR in dB.

**Fig. 7**. Illustration of WERs in % and fwSegSNR levels in dB, obtained at the input and at the output of the free-field- and HRTF-based beamformers.

## 4. SUMMARY AND CONCLUSION

In this paper, an HRTF-based beamformer design was presented which incorporates HRTFs instead of free-field steering vectors into its optimization problem. It therefore can account for effects of the head of a humanoid robot on the sound field. Experiments have shown that the signal enhancement performance of the new HRTF-based beamformer is significantly higher than of the original free-field design. Future work includes a detailed analysis of the proposed HRTF-based beamformer design with respect to various head orientations relative to the robot's body, with respect to the robustness against estimation errors of the DoA of the desired source, and with respect to the signal enhancement performance in different acoustic environments. Moreover, the HRTF-based design will be extended to the concept of polynomial beamforming [15], to allow for a flexible steering of the beamformer.

## 5. REFERENCES


[1] M. Pedersen, U. Kjems, K. Rasmussen, and L. Hansen, "Semi-blind source separation using head-related transfer functions," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5, May 2004, pp. V–713–16.

[2] M. Maazaoui, K. Abed-Meraim, and Y. Grenier, "Blind source separation for robot audition using fixed hrtf beamforming," *EURASIP Journal on Advances in Signal Processing (JASP)*, vol. 2012, p. 58, 2012.

[3] ——, "Adaptive blind source separation with hrtfs beamforming preprocessing," in *IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, June 2012, pp. 269–272.

[4] M. Maazaoui, Y. Grenier, and K. Abed-Meraim, "From binaural to multimicrophone blind source separation using fixed beamforming with hrtfs," in *IEEE Int. Conf. on Systems, Signals and Image Processing (IWSSIP)*, April 2012, pp. 480–483.

[5] E. Mabande, A. Schad, and W. Kellermann, "Design of robust superdirective beamformers as a convex optimization problem," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, April 2009, pp. 77–80.

[6] W. Herbordt, *Sound Capture for Human/Machine Interfaces - Practical Aspects of Microphone Array Signal Processing*. Heidelberg, Germany: Springer, March 2005.

[7] H. Van Trees, *Detection, Estimation, and Modulation Theory, Optimum Array Processing*, ser. Detection, Estimation, and Modulation Theory. Wiley, 2004.

[8] M. R. Schroeder, "Integrated–impulse method measuring sound decay without using impulses," *The Journal of the Acoustical Society of America (JASA)*, vol. 66, no. 2, pp. 497–500, 1979.

[9] D. Huggins-Daines, M. Kumar, A. Chan, A. Black, M. Ravishankar, and A. Rudnicky, "Pocketsphinx: A free, real-time continuous speech recognition system for hand-held devices," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2006, pp. I–183–88.

[10] M. Cooke, J. Barker, S. Cunningham, and X. Shao, "An audio-visual corpus for speech perception and automatic speech recognition," *The Journal of the Acoustical Society of America (JASA)*, vol. 120, no. 5, pp. 2421–2424, November 2006.

[11] H. Christensen, J. Barker, N. Ma, and P. Green, "The CHiME corpus: A resource and a challenge for computational hearing in multisource environments." in *INTERSPEECH*. International Speech Communication Association (ISCA), September 2010, pp. 1918–1921.

[12] Y. Hu and P. Loizou, "Evaluation of objective quality measures for speech enhancement," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 16, no. 1, pp. 229–238, January 2008.

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

[14] *NAO NEXT Gen H25 Datasheet*, Aldebaran Robotics, December 2011.

[15] E. Mabande and W. Kellermann, "Design of robust polynomial beamformers as a convex optimization problem," in *IEEE Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, August 2010.