2013 IEEE INTERNATIONAL WORKSHOP ON MACHINE LEARNING FOR SIGNAL PROCESSING, SEPT. 22–25, 2013, SOUTHAMPTON, UK

INFORMED GENERALIZED SIDELOBE CANCELER UTILIZING SPARSITY OF SPEECH SIGNALS

Jiří Málek[1], *Zbyněk Koldovský*[1,3], *Sharon Gannot*[2] *and Petr Tichavský*[3]

[1] Faculty of Mechatronics, Informatics, and Interdisciplinary Studies,
Technical University of Liberec, Studentská 2, 461 17, Liberec, Czech Republic.
E-mail:{jiri.malek,zbynek.koldovsky}@tul.cz,
[2] Faculty of Engineering, Bar-Ilan University, Ramat-Gan 52900, Israel.
[3]Institute of Information Theory and Automation, Pod vodárenskou věží 4,
P.O. Box 18, 182 08, Praha 8, Czech Republic.

ABSTRACT

This report proposes a novel variant of the generalized sidelobe canceler. It assumes that a set of prepared relative transfer functions (RTFs) is available for several potential positions of a target source within a confined area. The key problem here is to select the correct RTF at any time, even when the exact position of the target is unknown and interfering sources are present. We propose to select the RTF based on ℓ_p-norm, $p \leq 1$, measured at the blocking matrix output in the frequency domain. Subsequent experiments show that this approach significantly outperforms previously proposed methods for selection when the target and interferer signals are speech signals.

Index Terms— Noise Extraction; Speech Enhancement; ℓ_p-norm Minimization; Generalized Sidelobe Canceler; Semi-Blind Source Separation

1. INTRODUCTION

Generalized Sidelobe Canceler (GSC) is a popular implementation of the Minimum Variance Distortionless Response (MVDR) beamformer used in audio signal processing. It is comprised of three blocks, called *fixed beamformer* (FB), *blocking matrix* (BM), and *adaptive interference canceler* (AIC). The FB block acquires the target signal under the distortionless constraint. The BM works in parallel with the FB. Its purpose is to cancel the target signal and produce noise-only reference signals. The AIC cancels the residual noise at the output of the FB, given the output of the BM.

The original GSC [1] assumes anechoic propagation of the sound. It therefore fails in real-world environments that are highly reverberant. Gannot et al. [2] proposed a variant called Transfer Function-GSC (TF-GSC) that relies on estimated relative transfer functions (RTFs) between the target source and the microphones. The goal of TF-GSC is to retrieve the responses of the target on a reference microphone (dereverberation is not the goal). The beamformer is also typically endowed by a post-filtering block that suppresses residual noise, using approaches inspired by methods designed for single-channel signal enhancement [3, 4].

The performance of TF-GSC depends strongly on its knowledge of the RTFs, which depend on the position of the target and can be measured during periods when only the target is active; however, once the target moves, new RTFs must be obtained. Otherwise, the target signal leaks through the BM, which causes a distortion in the output of the beamformer. The key problem is to determine the RTFs when noise is present and the position of the target is not known.

Various solutions have been proposed [6, 5] including those using blind source separation [7]. Recently, we have focused on a solution that relies on knowledge of a set of prepared RTFs for several potential positions of the target [9]. This is useful in situations where the target's position is limited to a particular area, and the set of RTFs is prepared or progressively completed/updated during target-only intervals. In this paper, we will assume that the set is already prepared. The main focus is on the selection of proper RTFs from the set when noise is present.

In [10], a simplistic solution called the minimum variance approach (MVA) is used. MVA selects the RTFs for which the output variance of the BM is minimum. It relies on the assumption that a correctly canceled target at the BM output yields reduced signal power; however, this is not generally the case. MVA often fails to select the correct RTFs, particularly when the signal-to-noise ratio (SNR) is below 0 dB. In [8, 9], an approach, from here denoted as ICA, was proposed. It uses independent component analysis to find an optimum linear combination of the known RTFs, which is more flexible

This work was supported by Grant Agency of the Czech Republic through the project P103/11/1947.

978-1-4799-1180-6/13/$31.00

than the pure selection of RTFs. The combination is searched so that the output of the BM is as independent as possible. ICA was shown to improve the blocking ability of the BM compared to MVA, especially for an SNR below 0 dB.

In this paper, we propose to select the RTFs (not a linear combination) based on ℓ_p-norm, $p \leq 1$, measuring the signals' sparsity at the BM output in the frequency domain. For speech signals, which are sparse in that domain, the approach significantly outperforms MVA and ICA in terms of correctly selected RTFs. The complexity of this approach is comparable with MVA, which makes it available for systems with many microphones. Based on this new approach, we propose a modified TF-GSC beamformer, in which the RTFs are changed according to the position of the target speaker. We present several experiments with a fixed as well as moving target speaker, and one interfering speaker. We also present results of the SiSEC 2013 evaluation campaign[1].

2. NOTATIONS

A signal $s(n)$ of a directional (possibly moving) target source observed on the mth microphone, $m = 1, \ldots, M$, is described by

$$x_m(n) = \{h_{m,n} * s\}(n) + y_m(n) \tag{1}$$

where $h_{m,n}$ denotes the room impulse response between the target and the mth microphone, n is the time index, $*$ denotes the convolution, and $y_m(n)$ denotes unwanted signals from other sources, commonly referred to as interference. An approximate description in the short-time Fourier transform (STFT) domain reads

$$X_m(t,k) = H_m(t,k)S(t,k) + Y_m(t,k), \tag{2}$$

where t is the frame index, and k is the frequency index. Here, the dependency of $h_{m,n}$ on n is embodied by the dependency of $H_m(t,k)$ on t. In fact, it is assumed that $h_{m,n}$ is changing slowly and is approximately constant within each frame.

In a vector notation, (2) can be written as

$$\mathbf{X}(t,k) = \mathbf{H}(t,k) \cdot S(t,k) + \mathbf{Y}(t,k), \tag{3}$$

where

$$\begin{aligned} \mathbf{X}^T(t,k) &= [X_1(t,k)\ X_2(t,k)\ \ldots\ X_M(t,k)], \\ \mathbf{H}^T(t,k) &= [H_1(t,k)\ H_2(t,k)\ \ldots\ H_M(t,k)], \\ \mathbf{Y}^T(t,k) &= [Y_1(t,k)\ Y_2(t,k)\ \ldots\ Y_M(t,k)]. \end{aligned} \tag{4}$$

Next, we can write (3) as

$$\mathbf{X}(t,k) = \mathbf{A}(t,k) \cdot S_1(t,k) + \mathbf{Y}(t,k), \tag{5}$$

where

$$\mathbf{A}^T(t,k) = \left[1\ \frac{H_2(t,k)}{H_1(t,k)}\ \ldots\ \frac{H_M(t,k)}{H_1(t,k)}\right] \tag{6}$$

[1] http://sisec.wiki.irisa.fr

is the vector of RTFs, and $S_1(t,k) = H_1(t,k)S(t,k)$ is the response of $S(t,k)$ on the first microphone, which is the target signal. Without much loss on generality, we may assume that $H_1(t,k) \neq 0$.

3. INFORMED BEAMFORMING

As stated in the introduction, we assume that RTFs are known for several potential positions of a target speaker that is located within a limited area. Let the known RTFs be denoted by $\mathbf{A}_i(k)$, $i = 1, \ldots, I$. Our goal is to derive a method that selects the best fitting $\mathbf{A}_i(k)$ and performs beamforming to yield the best possible estimate of $S_1(t,k)$. The method is based on Transfer Function-GSC (TF-GSC) from [2], which is an efficient beamformer provided that $\mathbf{A}(t,k)$ is known. TF-GSC is described briefly in the following subsection.

3.1. Transfer Function-Generalized Sidelobe Canceler

In TF-GSC, the fixed beamformer is represented by $\mathbf{W}(t,k)$, and its output is

$$S_{FB}(t,k) = \mathbf{W}^H(t,k)\mathbf{X}(t,k) \tag{7}$$

where ^{H} denotes the conjugate transpose. The choice in [2] is $\mathbf{W}(t,k) = \mathbf{A}(t,k)$, where $\mathbf{A}(t,k)$ is assumed to be known (or estimated), satisfies the distortion-less constraint up to the scale factor $\|\mathbf{A}(t,k)\|^2$, which is, nevertheless, approximately constant.

The blocking matrix $\mathcal{B}(t,k)$ is an $M \times (M-1)$ MIMO filter defined as

$$\mathcal{B}(t,k) = \begin{bmatrix} -W_2^*(t,k) & -W_3^*(t,k) & \ldots & -W_M^*(t,k) \\ 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ & \ldots & \ddots & \\ 0 & 0 & \ldots & 1 \end{bmatrix}, \tag{8}$$

where * denotes complex conjugation. It is required that the columns of $\mathcal{B}(t,k)$ must be orthogonal to $\mathbf{A}(t,k)$ so that the output yields $M-1$ noise-only reference signals $U_m(t,k)$, $m = 1, \ldots, M-1$, given as elements of

$$\mathbf{U}(t,k) = \mathcal{B}^H(t,k)\mathbf{X}(t,k). \tag{9}$$

The adaptive interference canceler is represented by $(M-1) \times 1$ vector $\mathbf{G}(t,k)$ that estimates the residual noise in $S_{FB}(t,k)$ as $\mathbf{G}^H(t,k)\mathbf{U}(t,k)$. The output of the entire beamformer is then

$$\hat{S}(t,k) = S_{FB}(t,k) - \mathbf{G}^H(t,k)\mathbf{U}(t,k). \tag{10}$$

$\mathbf{G}(t,k)$ can be searched adaptively by minimizing

$$E\{\left\|S_{FB}(t,k) - \mathbf{G}^H(t,k)\mathbf{U}(t,k)\right\|^2\}, \tag{11}$$

where $E\{\cdot\}$ denotes the expectation operator, at times when noise signals are dominant. In TF-GSC, an adaptive normalized least mean square (NLMS) algorithm is used. The filter is updated according to

$$\mathbf{G}(t,k) \leftarrow \mathbf{G}(t,k) + \nu \frac{\mathbf{U}(t,k)\hat{S}(t,k)}{P(t,k)}, \qquad (12)$$

and is then truncated in the time-domain to avoid the effect of circular convolution. Here, $\nu \in (0,2)$ is a step-size parameter, and $P(t,k)$ is the average power-spectrum of the noise reference signals, i.e.

$$P(t,k) = \|\mathbf{U}(t,k)\|^2 / (M-1). \qquad (13)$$

The authors of [2] also suggest normalizing the update in (12) by the norm of the input signals, $\mathbf{X}(t,k)$.

3.2. Choice of the RTF

Let the blocking matrix, defined according to (8) where $\mathbf{W}(t,k) = \mathbf{A}_\alpha(k)$, be denoted by $\mathcal{B}_\alpha(k)$, $\alpha = 1,\ldots,I$. In this paper, we propose to select α such that the output $\mathbf{U}^\alpha(t,k) = \mathcal{B}_\alpha^H(k)\mathbf{X}(t,k)$ yields minimum ℓ_p-norm, $p \leq 1$, namely,

$$\alpha = \arg \min_{j\in\{1,\ldots,I\}} \left(\sum_{m=1}^{M-1} \sum_k |U_m^j(t,k)|^p \right)^{\frac{1}{p}}, \qquad (14)$$

where $U_m^j(t,k)$ is the mth element of $\mathbf{U}^j(t,k)$. Note that for $p = 2$ the criterion corresponds to the variance, and the proposed method is equivalent to the MVA. The following reasons lead us to the choice of $p \in (0,1]$.

- It may be expected, especially when both the target and interference are speech signals, that the correctly performing BM yields signals with sparser spectra, since it cancels one speech signal. The fact that the ℓ_p-norm is measured in the short-term frequency domain is crucial, because speech signals are known to be sparse in this domain.
- The ℓ_p-norm is sparsity enforcing criterion for $p \leq 1$.
- For SNR above 0 dB, the output variance still contains valuable information [10]. It is therefore handy that the criterion in (14) is proportional to the scale of the signal.

3.3. TF-GSC with changing RTF

Now we propose a modified TF-GSC beamformer that utilizes the set of available RTFs; for clarity, the method will be abbreviated IGSC (Informed GSC). The beamformer starts processing each frame by computing α according to (14). The FB and BM parts are then selected, respectively, as $\mathbf{W}(t,k) = \mathbf{A}_\alpha(k)$ and $\mathcal{B}(t,k) = \mathcal{B}_\alpha(k)$.

The AIC part of this beamformer is performed by selecting $\mathbf{G}(t,k) = \mathbf{G}_\alpha(t,k)$ where vectors $\mathbf{G}_1(t,k),\ldots,\mathbf{G}_I(t,k)$ are stored in memory. For the given frame, only $\mathbf{G}_\alpha(t,k)$ is updated by the NLMS algorithm, while the other vectors are kept the same. This helps the AIC to quickly adapt to the changing $\mathbf{A}(t,k)$.

The beamformer is endowed by a Wiener-like post-filter applied to the output signal $\hat{S}(t,k)$. The filter is defined through

$$V(t,k) = \frac{|\hat{S}(t,k)|^2}{|\hat{S}(t,k)|^2 + |\mathbf{G}_\alpha^H(t,k)\mathbf{U}^\alpha(t,k)|^2}. \qquad (15)$$

To summarize, processing of a frame proceeds as follows:

1. For each $j = 1,\ldots,I$, compute $\mathbf{U}^j(t,k) = \mathcal{B}_j^H(k)\mathbf{X}(t,k)$.
2. Find α according to (14) and put $\mathbf{W}(t,k) = \mathbf{A}_\alpha(k)$, $\mathcal{B}(t,k) = \mathcal{B}_\alpha(k)$, and $\mathbf{G}(t,k) = \mathbf{G}_\alpha(t,k)$.
3. Compute the beamformer's output as $\hat{S}(t,k) = S_{FB}(t,k) - \mathbf{G}_\alpha^H(t,k)\mathbf{U}^\alpha(t,k)$.
4. Update $\mathbf{G}_\alpha(t,k)$ according to (12) and (13).
5. Apply post-filter (15), so the final output is $\tilde{S}(t,k) = V(t,k)\hat{S}(t,k)$.

3.4. Preparation of RTFs

RTFs for a particular position of the target source can be estimated from noise-free recordings of the source [2, 9, 12, 13][2]. We assume that the target's location is confined to a certain area and prepare the set of RTFs for I positions that are regularly distributed within that area.

To this end, a noise-free recording is obtained for each such position by playing a speech signal from a loudspeaker placed at that position. RTFs are then estimated in the time domain. Let $z_m^i(n)$ denote the noise-free recording for the ith position obtained by the mth microphone. The impulse response of the RTF, for the mth microphone, is obtained as the solution of the least-squares problem [9]

$$a_m^i = \arg\min_a \sum_{n=1}^{N} \left| \{a * z_1^i\}(n) - z_m^i(n-d) \right|^2 \qquad (16)$$

where N is the sample length of $z_m^i(n)$, and d is a short integer delay introduced due to causality. The mth element of $\mathbf{A}_i(k)$ is then obtained as the kth element of the Fourier transform of a_m^i.

[2] An unbiased estimation of the RTF is possible also when stationary noise is present [2].

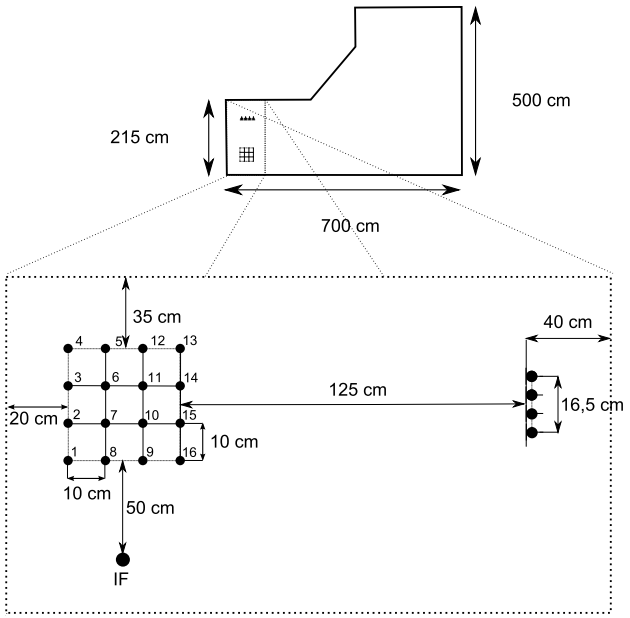

Fig. 1. Illustration of the experimental setup in our (non-rectangular) office. The known positions of the target source, located in a regular grid, are numbered 1 through 16.

4. EXPERIMENTAL EVALUATION

The experiments presented in this article were conducted in an office depicted in Fig. 1 to verify the efficiency of the criterion (14) and to evaluate the performance of the proposed IGSC. Two scenarios are considered, respectively, where the location of the target speaker is fixed or moving within a 30×30 cm area whose center was at a 1.25 m distance from a linear microphone array with $M = 4$ microphones. The interference is another speaker located in the fixed position denoted by IF in Fig. 1. The reverberation time T_{60} of the room is about 490 ms, the inter-microphone distance is 5.5 cm, and the sampling frequency is 16 kHz.

The RTFs were computed for the positions within the area using 4 s of a female interference-free utterance played from each position. For testing, the target and interference signals consist of 4 s of female and 8 s of male utterances, uttered by different speakers. The speech signals were taken from the TIMIT database [15].

4.1. Fixed target position

The aim of this experiment is to analyze the percentage of the correctly determined RTF, using various approaches specified below. The target source is located consecutively in all positions of the grid (Fig. 1). The correct RTF is therefore known as a ground truth. Next, we also evaluate the Interference-to-Signal Ratio (ISR) at the output of the BM, which reflects the degree of the target cancellation.

We compare the following methods that are all applied in the frequency domain where the length of the STFT frame was 2048 samples with a shift of 128 samples[3].

1. The *minimum variance approach* (MVA) [10].
2. The kurtosis-based approach[4] that selects the RTF exhibiting maximum sample kurtosis at the BM output. In place of (14), it computes
$$\alpha = \arg\min_j \sum_{m=1}^{M-1} \left| \frac{\sum_k |U_m^j(t,k)|^4}{(\sum_k |U_m^j(t,k)|^2)^2} - 3 \right|. \quad (17)$$
3. The method based on ICA derived in [9].
4. An "oracle" approach denoted by trueRTF that selects the ground truth RTF.
5. Another "oracle" approach, denoted by bestRTF, that selects the RTF for which the BM output yields the maximum Interference-to-Signal Ratio (SIR). This approach is designed for the moving target scenario in the next subsection, where the ground truth RTF is not given.

The experiment was repeated for each position. The average percentage is shown in Table 1 and the average ISR at the BM output is shown in Fig. 2.

	input Signal-to-Interference Ratio				
	−10 dB	−5 dB	0 dB	5 dB	10 dB
ℓ_1 norm	62.76	80.53	91.24	96.21	97.77
$\ell_{0.5}$ norm	**74.27**	**87.52**	**94.04**	**96.79**	**97.96**
$\ell_{0.1}$ norm	72.82	85.23	91.85	94.97	96.45
ℓ_1 norm (TD)	29.63	49.35	67.90	80.44	87.55
$\ell_{0.5}$ norm (TD)	27.80	47.09	65.84	78.95	86.31
ICA	58.70	75.13	85.58	91.64	94.43
MVA	28.33	47.70	67.17	80.29	88.01
kurtosis	15.23	19.97	21.73	19.60	17.47

Table 1. Results in terms of correctly determined RTF [%].

The proposed ℓ_p-norm-based approach achieves the best results, especially for $p \approx 0.5$. Its performance significantly drops when (14) is evaluated on signals in the time domain, denoted by TD in Table 1, which is not the sparsity domain of speech signals. The ICA-based approach from [9] achieves the second-best performance, however, at a much higher computational burden.

MVA performs the same in the time domain as in the frequency domain, due to the Parseval equality. Therefore, it gives similar results to the ℓ_p-norm computed in the TD. The kurtosis-based approach failed in this experiment. The main reason seems to be the fact that (17) is invariant to the scale

[3] In total, there are 1485 frames to analyze.

[4] Kurtosis is used as a contrast function in some ICA methods to separate independent sources by maximizing their non-Gaussianity [14].

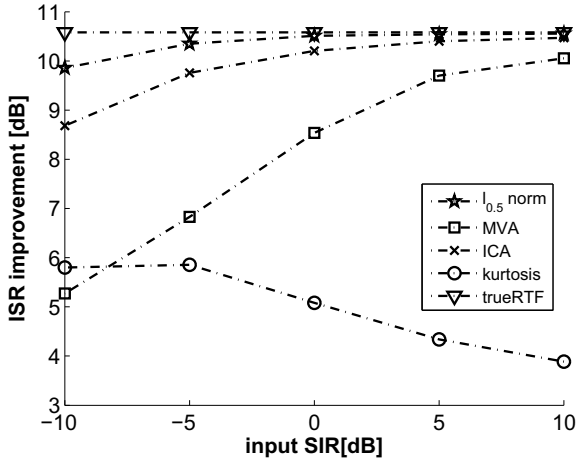

Fig. 2. The average ISR improvement at the BM output achieved by the compared approaches when the target's location is fixed.

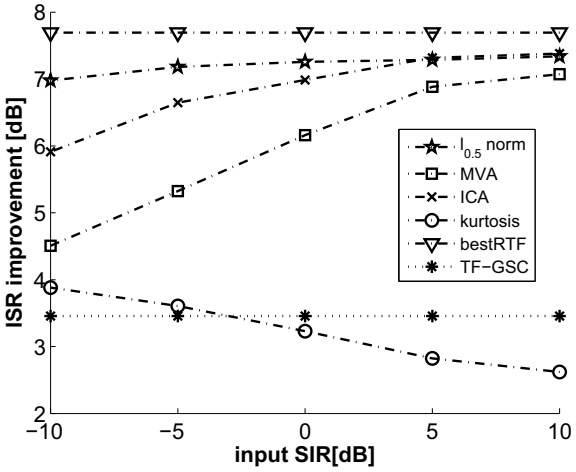

Fig. 3. ISR improvement averaged over two examples with moving target speaker.

of signals. This is interesting in view of another fact that the scale of input signals is also not relevant for the ICA-based approach, although the ICA approach performs well.

4.2. Moving target scenario

In this experiment, two noisy recordings were considered, of a target that moves across positions 1 through 9 and 16 through 9, respectively. The proposed IGSC with $p = 0.5$ is applied to enhance the target signal. Its performance is compared with that of the original TF-GSC beamformer [2] and with IGSC which is endowed by the optimal "bestRTF" selection procedure. TF-GSC assumes fixed RTFs, so we choose suboptimal RTFs for position 7 and 10, respectively, for the two movements.

Fig. 3 shows the average ISR improvement at the outputs of BMs. In this scenario, the target occurs in positions for which the exact RTFs are not available (between the points). Therefore, the results are generally lower by about 3 dB compared to Fig. 2. The "bestRTF" procedure yields the maximum attainable performance, which is best approached by the $\ell_{0.5}$-norm-based method. The BM of TF-GSC loses about 3.5 dB uniformly, due to the fixed RTFs.

The enhancement of the target signal by the compared beamformers is evaluated in terms of Signal-to-Interference (SIR) and Signal-to-Distortion (SDR) ratios[5] in Figures 4 and 5, respectively. These results are consistent with those in Fig. 3. In other words, the SIR and SDR of enhanced signals depend on the blocking ability of the BM part of beamformers. The proposed IGSC yields better enhancement of the target signal than the original TF-GSC. Its performance is tight to the performance of IGSC using "bestRTF" even when the input SIR goes below 0 dB.

[5]The criteria are defined as in [9].

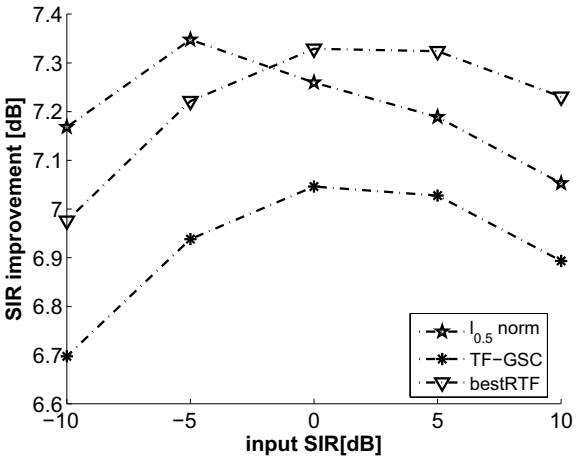

Fig. 4. SIR improvement achieved by compared beamformers as a function of input SIR.

4.3. SiSEC 2013

A modified version of IGSC takes part in the SiSEC 2013 evaluation campaign in the task titled "Two-channel noisy recordings of a moving speaker within a limited area". In this task, the target is a loudspeaker that occurs within a 30x30cm area. It is recorded by two microphones that are 2 meters distant from the center of the area. A development dataset is provided that contains noise-free recordings of the target from 16 positions within the area, which we use for the preparation of the set of RTFs for IGSC. A testing dataset contains recordings of the target, which performs movements within the area, and an omnidirectional babble noise.

The modifications of IGSC are as follows. The AIC part is only used to adapt filters $\mathbf{G}_1(t,k),\ldots,\mathbf{G}_I(t,k)$. The interference cancellation (step 3 in Section 3.3) is not performed,

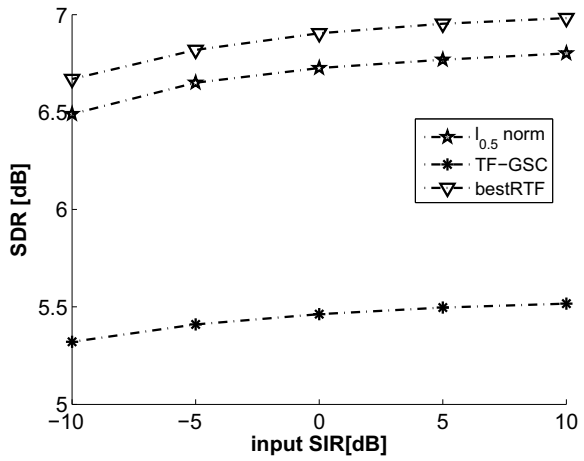

Fig. 5. SDR achieved by compared beamformers as a function of input SIR.

because the noise is not directional. Only the post-filtering is done to attenuate the residual noise at the FB output. Next, we use the post-filter by [16] instead of the Wiener filter (15), because it yields better perceptual quality of the estimated signal. The results are reported on the internet site of the SiSEC 2013 campaign (`http://sisec.wiki.irisa.fr`).

5. CONCLUSIONS

The selection of RTFs from a prepared set has been improved using the ℓ_p-norm, especially when the power of the target signal is the same or lower than that of the interference. Using this approach, a novel variant of informed GSC has been proposed. The beamformer is able to enhance a target signal coming from the assumed area better than TF-GSC, assuming a fixed position of the target.

6. REFERENCES

[1] L. Griffiths and C. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propag.*, vol. 30, no. 1, pp. 27–34, Jan. 1982.

[2] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. on Signal Processing*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.

[3] R. Zelinski, "A microphone array with adaptive postfiltering for noise reduction in reverberant rooms," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Munich, Germany, pp. 2578–2581, 1988.

[4] S. Gannot and I. Cohen, "Speech Enhancement Based on the General Transfer Function GSC and Postfiltering," *IEEE Trans. on Speech and Audio Processing*, vol. 12, no. 6, pp. 561–571, Nov. 2004.

[5] O. Hoshuyama, A. Sugiyama, A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," *IEEE Transactions on Signal Processing*, vol.47, no. 10, pp. 2677–2684, Oct. 1999.

[6] W. Herbordt, W. Kellermann, "Analysis of blocking matrices for generalized sidelobe cancellers for non-stationary broadband signals," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2002)*, vol. 4, pp. IV-4187, May 2002.

[7] S. Makino, Te-Won Lee, and H. Sawada, *Blind Speech Separation*, Springer, Sept. 2007.

[8] J. Málek, Z. Koldovský and P. Tichavský, "Semi-Blind Source Separation Based on ICA and Overlapped Speech Detection", *Proc. of The 10th International Conference on Latent Variable Analysis and Source Separation (LVA/ICA 2012)*, LNCS 7191, pp. 462-469, Tel-Aviv, Israel, March 12-15, 2012.

[9] Z. Koldovský, J. Málek, P. Tichavský, and F. Nesta, "Semi-blind Noise Extraction Using Partially Known Position of the Target Source", accepted in *IEEE Trans. on Speech, Audio and Language Processing*, 2013.

[10] Z. Koldovský, P. Tichavský, D. Botka, "Noise Reduction in Dual-Microphone Mobile Phones Using A Bank of Pre-Measured Target-Cancellation Filters," *Proc. of ICASSP 2013*, pp. 679–683, Vancouver, Canada, May 2013.

[11] P. Smaragdis, "Position and trajectory learning for microphone arrays," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 15, no. 1, pp. 358–368, Jan. 2007.

[12] A. Krueger, E. Warsitz, and R. Haeb-Umbach, "Speech enhancement with a GSC-like structure employing eigenvector-based transfer function ratios estimation," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 19, no. 1, Jan. 2011.

[13] R. Talmon, I. Cohen, and S. Gannot, "Relative transfer function identification using convolutive transfer function approximation," , *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 17, no. 4, pp. 546–555, May 2009.

[14] A. Hyvärinen and E. Oja, "A Fast Fixed-Point Algorithm for Independent Component Analysis". *Neural Computation*, vol. 9, no. 7, pp. 1483–1492, 1997.

[15] J. S. Garofolo, et al., "TIMIT Acoustic-Phonetic Continuous Speech Corpus", Linguistic Data Consortium, Philadelphia, 1993.

[16] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, no. 2, pp.443–445, Apr. 1985.