DUAL SOURCE TRANSFER-FUNCTION GENERALIZED SIDELOBE CANCELLER

Gal Reuven[1], *Sharon Gannot*[2] *and Israel Cohen*[1]

[1]Faculty of Electrical Engineering, Technion - Israel Institute of Technology, Haifa 32000, Israel
[2]School of Engineering, Bar-Ilan University, Ramat-Gan, 52900, Israel
galrv@tx.technion.ac.il;gannot@eng.biu.ac.il; icohen@ee.technion.ac.il

ABSTRACT

Full duplex hands-free man/machine interface often suffers from directional non-stationary interference (such as a competing speaker or an echo signal) as well as a stationary interference (which may comprise both directional and non-directional signals). We propose a new structure for handling both interferences, based on the *transfer-function generalized sidelobe canceller* (TF-GSC). As in the standard GSC structure, it contains three blocks: a *matched-TF beamformer* (MTFBF), a *blocking matrix* and a multi-channel *noise canceller*. The blocking matrix is designed to block both the desired and the competing speech signals. The MTFBF is designed to block the competing speech signal, while maintaining the desired signal arriving from the direction of interest. The multi-channel noise canceller is used to mitigate the remaining noise in the MTFBF output. The performance of the proposed algorithm is evaluated through a series of simulations, in reverberant environments. This structure is shown to be related to both the *echo cancellation* problem and to the convolutive *blind source separation* (BSS) problem.

1. INTRODUCTION

Signal quality might significantly deteriorate in the presence of interferences. The *transfer-function generalized sidelobe canceller* (TF-GSC) proposed in [1] uses beamforming and non-stationarity to enhance speech signal deteriorated by a single stationary interference signal in an arbitrary transfer function (TF) enclosure. When a second, directional, non-stationary interference signal is also present, the algorithm is unable to distinguish between the desired signal and the interference and therefore a performance degradation is unavoidable.

This problem may be viewed as a convolutive *blind source separation* (BSS) problem, where the desired and the competing speech signals are filtered by the room impulse response before being mixed. The signals received are also contaminated by additive noise sources. Speech signal non-stationarity is exploited by Parra and Spence [2] to obtain a nonlinear minimization problem. Both permutation and gain ambiguity problems, encountered by the frequency domain approach are alleviated by imposing an FIR structure on the mixing filters.

Our problem is closely related to echo cancellation problem as well. In these problems a joint effort of mitigating the echo signal and reducing the noise level is required. However, the two tasks generally contradict each other [3], especially in situations where both signals are active (usually, denoted a *double talk* situation). However, it should be stressed that in the echo cancellation problem a separate measurement of the interference signal is available and can be used to improve the performance of the overall system.

Affes and Grenier proposed in [4] a GSC structure for double talk situations. They presented a distortionless fixed beamformer constrained to cancel the echo, and a blocking matrix constrained to block both the desired signal and echo signal. The TFs are estimated using subspace tracking methods. These estimates are used to construct both the fixed beamformer and the blocking matrix.

In [3] two frequency domain schemes for joint echo cancellation and noise reduction are presented. Both contain the TF-GSC beamformer proposed in [1] and a block *least mean square* (LMS) *acoustic echo canceller* (AEC). Following Kellermann [5], the first scheme comprises multi-channel AEC followed by a beamformer, while the second comprises a beamformer followed by a single channel AEC as a post-filter. A series of simulations using real speech recordings showed that the first scheme outperforms the second one. Two additional schemes for noise reduction and echo cancellation are proposed and compared in [6]. The first scheme includes a multi-channel AEC followed by a *generalized singular value decomposition* (GSVD) based beamformer. The second scheme incorporates the far-end echo reference into the GSVD beamformer. Simulations indicate that the first scheme outperforms the second one.

In this paper we present a novel algorithm, based on the TF-GSC [1], for cancelling two interference signals, one stationary and the second non-stationary. Nevertheless, the algorithm is also applicable in scenarios where the non-stationary interference is an echo signal, if the available echo signal is properly incorporated into the scheme.

The structure of this work is as follows. In Sec. 2 we formulate the problem of dual-source interference cancelling in a general *acoustical transfer function* (ATF) environment. The proposed algorithm and its derivation are presented in Sec. 3. Sec. 4 deals with the ATF estimation procedure. Sec. 5 demonstrates some experimental results in practical scenarios. In Sec. 6 we discuss several applications of the proposed scheme.

2. PROBLEM FORMULATION

Let us consider an array of sensors in a dual source and noisy environment. The received signal includes three components, the desired speech source, the directional non-stationary interference signal (competing speech) and a stationary noise signal (which can be either directional or non-directional or a combination thereof). The mth microphone signal is

$$z_m(t) = a_m(t) * s_1(t) + b_m(t) * s_2(t) + n_m(t);\ m = 1, \dots, M \quad (1)$$

where $a_m(t)$ and $b_m(t)$ are the acoustical impulse responses relating the mth microphone and the desired speech source and the non-stationary interference, respectively. $s_1(t)$ and $s_2(t)$ are the desired speech source and the non-stationary interference source, respectively. $n_m(t)$ is the (directional or non-directional) stationary interference signal at the mth microphone. $*$ denotes convolution. No separate measurement of the desired signal and the interferences signals are available. In the *short time Fourier transform* (STFT) domain, in vector form, Eq. (1) can be approximately rewritten as:

$$\mathbf{Z}(t,e^{j\omega}) = \mathbf{A}(e^{j\omega})S_1(t,e^{j\omega}) + \mathbf{B}(e^{j\omega})S_2(t,e^{j\omega}) + \mathbf{N}(t,e^{j\omega}) \quad (2)$$

where,

$$\begin{aligned}
\mathbf{Z}^T(t,e^{j\omega}) &= \left[\, Z_1(t,e^{j\omega}) \;\; Z_2(t,e^{j\omega}) \;\; \cdots \;\; Z_M(t,e^{j\omega}) \,\right] \\
\mathbf{A}^T(e^{j\omega}) &= \left[\, A_1(e^{j\omega}) \;\; A_2(e^{j\omega}) \;\; \cdots \;\; A_M(e^{j\omega}) \,\right] \\
\mathbf{B}^T(e^{j\omega}) &= \left[\, B_1(e^{j\omega}) \;\; B_2(e^{j\omega}) \;\; \cdots \;\; B_M(e^{j\omega}) \,\right] \\
\mathbf{N}^T(t,e^{j\omega}) &= \left[\, N_1(t,e^{j\omega}) \;\; N_2(t,e^{j\omega}) \;\; \cdots \;\; N_M(t,e^{j\omega}) \,\right]
\end{aligned}$$

and $Z_m(t,e^{j\omega}), S_1(t,e^{j\omega}), S_2(t,e^{j\omega})$ and $N_m(t,e^{j\omega})$ are the STFT of the respective signals. $A_m(e^{j\omega})$ and $B_m(e^{j\omega})$ are the ATF from the desired speech source and competing speech source to the mth microphone, respectively, assumed hereinafter to be time invariant over the observation period.

3. THE PROPOSED ALGORITHM

Let $W^*(t,e^{j\omega})\;; m = 1,\dots,M$ be a set of M filters,

$$\mathbf{W}^\dagger(t,e^{j\omega}) = \left[\, W_1^*(t,e^{j\omega}) \;\; W_2^*(t,e^{j\omega}) \;\; \cdots \;\; W_M^*(t,e^{j\omega}) \,\right]$$

where * denotes conjugation and † denotes conjugation transpose. A beamformer is realized by $Y(t,e^{j\omega}) = \mathbf{W}^\dagger(t,e^{j\omega})\mathbf{Z}(t,e^{j\omega})$:

$$\begin{aligned}
Y(t,e^{j\omega}) &= \mathbf{W}^\dagger(t,e^{j\omega})\mathbf{A}(e^{j\omega})S_1(t,e^{j\omega}) \\
&+ \mathbf{W}^\dagger(t,e^{j\omega})\mathbf{B}(e^{j\omega})S_2(t,e^{j\omega}) + \mathbf{W}^\dagger(t,e^{j\omega})\mathbf{N}(t,e^{j\omega}) \\
&\triangleq Y_{s_1}(t,e^{j\omega}) + Y_{s_2}(t,e^{j\omega}) + Y_n(t,e^{j\omega}) \quad (3)
\end{aligned}$$

where $Y_{s_1}(t,e^{j\omega})$ is the desired speech part, $Y_{s_2}(t,e^{j\omega})$ is the competing speech part and $Y_n(t,e^{j\omega})$ is the stationary noise part. The output power is given by:

$$E\{Y(t,e^{j\omega})Y^*(t,e^{j\omega})\} = \mathbf{W}^\dagger(t,e^{j\omega})\Phi_{\mathbf{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})$$

where $\mathbf{\Phi}_{ZZ}(t,e^{j\omega}) = E\{\mathbf{Z}(t,e^{j\omega})\mathbf{Z}^\dagger(t,e^{j\omega})\}$. We want to minimize the output power subject to the following constraints:

$$\begin{aligned}
Y_{s_1}(t,e^{j\omega}) &= \mathbf{W}^\dagger(t,e^{j\omega})\mathbf{A}(e^{j\omega})S_1(t,e^{j\omega}) = \mathcal{F}^*(e^{j\omega})S_1(t,e^{j\omega}) \\
Y_{s_2}(t,e^{j\omega}) &= \mathbf{W}^\dagger(t,e^{j\omega})\mathbf{B}(e^{j\omega})S_2(t,e^{j\omega}) = 0 \quad (4)
\end{aligned}$$

where $\mathcal{F}^*(e^{j\omega})$ is a predefined filter response. We thus have the following minimization problem:

$$\begin{aligned}
&\min_{\mathbf{W}} \left\{\mathbf{W}^\dagger(t,e^{j\omega})\Phi_{\mathbf{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})\right\} \text{ subject to} \quad (5) \\
&\mathbf{W}^\dagger(t,e^{j\omega})\mathbf{A}(e^{j\omega}) = \mathcal{F}^*(t,e^{j\omega}) \text{ and } \mathbf{W}^\dagger(t,e^{j\omega})\mathbf{B}(e^{j\omega}) = 0
\end{aligned}$$

Solution to the problem depicted in (5), may be obtained by minimizing the complex Lagrangian

$$\begin{aligned}
\mathcal{L}(\mathbf{W}) &= \mathbf{W}^\dagger(t,e^{jw})\Phi_{ZZ}(t,e^{jw})\mathbf{W}(t,e^{jw}) \quad (6) \\
&+\lambda_1 \left[\mathbf{W}^\dagger(t,e^{jw})\mathbf{A}(e^{jw}) - \mathcal{F}^*(e^{jw})\right] \\
&+\lambda_1^* \left[\mathbf{A}^\dagger(e^{jw})\mathbf{W}(t,e^{jw}) - \mathcal{F}(e^{jw})\right] \\
&+\lambda_2\mathbf{W}^\dagger(t,e^{jw})\mathbf{B}(e^{jw}) + \lambda_2^*\mathbf{B}^\dagger(e^{jw})\mathbf{W}(t,e^{jw})
\end{aligned}$$

where λ_1 λ_2 are the Lagrange multipliers. Although a closed-form solution, $\mathbf{W}^{opt}(t,e^{j\omega})$, for the minimization problem may be obtained, it might be too cumbersome and would not have the ability to track changes in the environment. Hence we will replace the closed-form solution with an adaptive solution. Imposing both constraints on the steepest descent solution yields:

$$\begin{aligned}
&\mathbf{W}(t+1,e^{j\omega}) = \quad (7) \\
&P(e^{j\omega})\left[\mathbf{W}(t,e^{j\omega}) - \mu\mathbf{Z}(t,e^{j\omega})Y^*(t,e^{j\omega})\right] + \mathbf{F}(e^{j\omega})
\end{aligned}$$

where

$$\begin{aligned}
&\mathbf{P}(e^{j\omega}) = \mathbf{I} - \alpha^{-1}\cdot \quad (8) \\
&\left[\mathbf{A}(e^{j\omega})\mathbf{A}^\dagger(e^{j\omega})\left(\left\|\mathbf{B}(e^{j\omega})\right\|^2 I - \mathbf{B}(e^{j\omega})\mathbf{B}^\dagger(e^{j\omega})\right)\right. \\
&\left. + \mathbf{B}(e^{j\omega})\mathbf{B}^\dagger(e^{j\omega})\left(\left\|\mathbf{A}(e^{j\omega})\right\|^2 I - \mathbf{A}(e^{j\omega})\mathbf{A}^\dagger(e^{j\omega})\right)\right]
\end{aligned}$$

$$\begin{aligned}
&\mathbf{F}(e^{j\omega}) = \alpha^{-1}\cdot \\
&\left(\left\|\mathbf{B}(e^{j\omega})\right\|^2 I - \mathbf{B}(e^{j\omega})\mathbf{B}^\dagger(e^{j\omega})\right)\mathbf{A}(e^{j\omega})\mathcal{F}(e^{j\omega})
\end{aligned}$$

and

$$\alpha \triangleq \mathbf{A}^\dagger(e^{j\omega})\left[\left\|\mathbf{B}(e^{j\omega})\right\|^2 I - \mathbf{B}(e^{j\omega})\mathbf{B}^\dagger(e^{j\omega})\right]\mathbf{A}(e^{j\omega})$$

Now, following the rational of the TF-GSC, we can (uniquely) split $\mathbf{W}(t,e^{j\omega})$ into a sum of two vectors in mutually orthogonal subspaces. One in the direction of $\mathbf{F}(e^{j\omega})$ [that is perpendicular to $\mathbf{B}(e^{j\omega})$ but not to $\mathbf{A}(e^{j\omega})$]; and the other in the null space of $\left[\,\mathbf{A}(e^{j\omega})\middle|\mathbf{B}(e^{j\omega})\,\right]$. Hence,

$$\mathbf{W}(t,e^{j\omega}) = \mathbf{W}_0(t,e^{j\omega}) - \mathbf{V}(t,e^{j\omega}) \quad (9)$$

By the definition of the null-space,

$$\mathbf{V}(t,e^{j\omega}) = \mathcal{H}(e^{j\omega})\mathbf{G}(t,e^{j\omega}) \quad (10)$$

where $\mathcal{H}(e^{j\omega})$ is some $M \times (M-2)$ matrix, such that the columns of $\mathcal{H}(e^{j\omega})$ span the null space of $\left[\,\mathbf{A}(e^{j\omega})\middle|\mathbf{B}(e^{j\omega})\,\right]$. The vector $\mathbf{G}(t,e^{j\omega})$ is an $(M-2)\times 1$ vector of adjustable filters. By considerations similar to the geometrical interpretation of Frost's algorithm, $\mathbf{W}_0(t,e^{j\omega}) = \mathbf{F}(e^{j\omega})$. The idea is summarized in Fig. (1). As in [1], it can be shown that the problem can be separated into two branches. The first satisfies the constraint and the second conducts the minimization. The output signal $Y(t,e^{j\omega})$ is given by

$$Y(t,e^{j\omega}) = Y_{\text{MTFBF}}(t,e^{j\omega}) - Y_{\text{NC}}(t,e^{j\omega}) \quad (11)$$

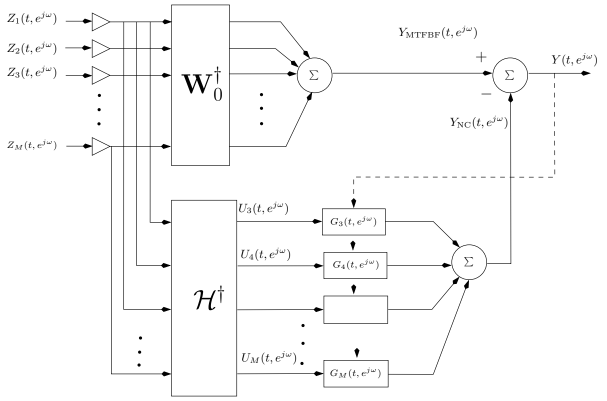

Figure 1: GSC solution for the dual source case.

where

$$
\begin{aligned}
Y_{\text{MTFBF}}(t, e^{j\omega}) &= \mathbf{W}_0^{\dagger}(t, e^{j\omega})\mathbf{Z}(t, e^{j\omega}) \\
Y_{\text{NC}}(t, e^{j\omega}) &= \mathbf{G}^{\dagger}(t, e^{j\omega})\mathcal{H}^{\dagger}(e^{j\omega})\mathbf{Z}(t, e^{j\omega})
\end{aligned} \tag{12}
$$

Namely, the solution comprises of three components: a MTFBF, a blocking matrix and a multichannel noise canceller. We now discuss each of these components.

3.1. Blocking Matrix

Consider the following $M \times (M-2)$ matrix $\mathcal{H}(e^{jw})$

$$
\mathcal{H}(e^{jw}) = \begin{bmatrix} h_3 & h_4 & \cdots & h_M \\ l_3 & l_4 & \cdots & l_M \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & \cdots & \ddots & \\ 0 & 0 & \cdots & 1 \end{bmatrix} \tag{13}
$$

where, for $m = 3, \ldots, M$

$$
h_m = -\frac{\frac{A_2^*}{A_1^*}\frac{B_m^*}{B_1^*} - \frac{B_2^*}{B_1^*}\frac{A_m^*}{A_1^*}}{\frac{A_2^*}{A_1^*} - \frac{B_2^*}{B_1^*}};\; l_m = -\frac{\frac{A_m^*}{A_1^*} - \frac{B_m^*}{B_1^*}}{\frac{A_2^*}{A_1^*} - \frac{B_2^*}{B_1^*}} \tag{14}
$$

It can be easily verified that $\mathcal{H}(e^{jw})$ satisfies

$$
\begin{cases} A^{\dagger}(e^{jw})\mathcal{H}(e^{jw}) = 0 \\ B^{\dagger}(e^{jw})\mathcal{H}(e^{jw}) = 0 \end{cases} \tag{15}
$$

and therefore is a valid blocking matrix. Define $\mathbf{U}(t, e^{j\omega}) = \mathcal{H}^{\dagger}(e^{j\omega})\mathbf{Z}(t, e^{j\omega})$, the reference noise signals. Therefore, using (13) we have for $m = 3, \ldots, M$:

$$
U_m(t, e^{j\omega}) = h_m(e^{j\omega})Z_1(t, e^{j\omega}) + l_m(e^{j\omega})Z_2(t, e^{j\omega}) + Z_m(t, e^{j\omega}) \tag{16}
$$

3.2. Matched Transfer Function Beamformer

The MTFBF is given by,

$$
\begin{aligned}
&\mathbf{W}_0(t, e^{j\omega}) = \mathbf{F}(t, e^{j\omega}) = \\
&\alpha^{-1}\left(\left\|\mathbf{B}(e^{j\omega})\right\|^2 I - \mathbf{B}(e^{j\omega})\mathbf{B}^{\dagger}(e^{j\omega})\right)\mathbf{A}(e^{j\omega})\mathcal{F}(e^{j\omega})
\end{aligned} \tag{17}
$$

Setting $\mathbf{W}_0(t, e^{j\omega})$ according to (17) assures that the constraints in (4) are satisfied.

3.3. Multi-Channel Noise Canceller

Minimization of the output power is obtained by adjusting the filters $G_m(t, e^{j\omega});\; m = 3, \ldots, M$. This minimization can be seen as the classical Widrow problem. Therefore, the filters can be obtained by:

$$
\begin{aligned}
\tilde{G}_m(t+1, e^{j\omega}) &= G_m(t, e^{j\omega}) + \mu \frac{U_m(t, e^{j\omega})Y^*(t, e^{j\omega})}{P_{\text{est}}(t, e^{j\omega})} \\
G_m(t+1, e^{j\omega}) &\overset{\text{FIR}}{\longleftarrow} \tilde{G}_m(t+1, e^{j\omega})
\end{aligned} \tag{18}
$$

where $P_{\text{est}}(t, e^{j\omega})$ is a power estimate used for normalization. The operator $\overset{\text{FIR}}{\longleftarrow}$ applies an FIR constraint in the time domain.

4. ATF ESTIMATION

Till this point, the ATFs were assumed to be known. However, in practice, they should be estimated. We assume that the ATFs ratios $\frac{\mathbf{A}(e^{j\omega})}{A_1(e^{j\omega})}$,and $\frac{\mathbf{B}(e^{j\omega})}{B_1(e^{j\omega})}$ are slowly changing in time compared to time variations of the desired signal and the directional interference. We also assume that the statistics of the noise signal is slowly changing compared with the statistics of both the desired signal and directional interference.

4.1. MTFBF Estimate

Estimation of the MTFBF is done in two steps. First, the ATFs ratios $\frac{\mathbf{A}(e^{j\omega})}{A_1(e^{j\omega})}$ and $\frac{\mathbf{B}(e^{j\omega})}{B_1(e^{j\omega})}$ are estimated, using the system identification procedure described in [1]. This procedure requires the division of the observation period into frames such that the desired or the competing speech signals may be considered stationary during each k-th frame. Note, however, that the two ratios cannot be estimated simultaneously, and we must use frames in which both signals are not simultaneously active.
In the second step, $\mathbf{F}(t, e^{j\omega})$ is estimated using (17), assuming that the two ratios estimates are valid, although they were estimated in distinct time periods. It can be shown that by using the ATFs ratios rather than the real ATFs in (17), the desired signal component of $Y_{\text{MTFBF}}(t, e^{j\omega})$ is distorted by $A_1(e^{j\omega})$, namely $\mathbf{F}(e^{j\omega})^{\dagger}\mathbf{A}(e^{j\omega}) = A_1(e^{j\omega})\mathcal{F}^*(e^{j\omega})$.

4.2. Blocking Matrix Estimate

Inspecting (14), we note that the filters $h_m(e^{jw})$ and $l_m(e^{jw})$ can be estimated by using the ATFs ratio estimates. Similar to $\mathbf{F}(t, e^{j\omega})$ estimation method this can be done in a two step procedure.
However, for the blocking matrix estimate we can use *double talk* situations to estimate $h_m(e^{jw})$ and $l_m(e^{jw})$ directly. Choose observation periods in which **both** the desired and competing speech signals are active simultaneously. Again, divide this period into frames such that speech signals may be considered stationary during each k-th frame. Using (16) we can obtain a system identification procedure,

$$
\begin{aligned}
&\Phi^{(k)}_{z_m z_1}(e^{j\omega}) = -h_m(e^{j\omega})\Phi^{(k)}_{z_1 z_1}(e^{j\omega}) \\
&\quad -l_m(e^{j\omega})\Phi^{(k)}_{z_2 z_1}(e^{j\omega}) + \Phi_{u_m z_1}(e^{j\omega});\; k = 1, \ldots, K
\end{aligned} \tag{19}
$$

where K is the number of frames in the interval, and $\Phi^{(k)}_{z_i z_j}(e^{j\omega})$ is the cross-PSD between z_i and z_j during the kth frame.

$\Phi_{u_m z_1}(e^{j\omega})$ is the cross-PSD between u_m and z_1. It is shown [1] that $\Phi_{u_m z_1}(e^{j\omega})$ is independent of the frame index k. By replacing real PSD values with their estimates, calculated using time-averages, the following vector equation is obtained:

$$\begin{bmatrix} \hat{\Phi}^{(1)}_{z_m z_1}(e^{j\omega}) \\ \hat{\Phi}^{(2)}_{z_m z_1}(e^{j\omega}) \\ \vdots \\ \hat{\Phi}^{(K)}_{z_m z_1}(e^{j\omega}) \end{bmatrix} = \begin{bmatrix} \hat{\Phi}^{(1)}_{z_1 z_1}(e^{j\omega}) & \hat{\Phi}^{(1)}_{z_2 z_1}(e^{j\omega}) & 1 \\ \hat{\Phi}^{(2)}_{z_1 z_1}(e^{j\omega}) & \hat{\Phi}^{(2)}_{z_2 z_1}(e^{j\omega}) & 1 \\ \vdots & & \\ \hat{\Phi}^{(K)}_{z_1 z_1}(e^{j\omega}) & \hat{\Phi}^{(K)}_{z_2 z_1}(e^{j\omega}) & 1 \end{bmatrix} \times \begin{bmatrix} -h_m(e^{j\omega}) \\ -l_m(e^{j\omega}) \\ \Phi_{u_m z_1}(e^{j\omega}) \end{bmatrix} + \begin{bmatrix} \varepsilon^{(1)}_m(e^{j\omega}) \\ \varepsilon^{(2)}_m(e^{j\omega}) \\ \vdots \\ \varepsilon^{(K)}_m(e^{j\omega}) \end{bmatrix} \tag{20}$$

(a separate set of equations is used for $m = 3, \ldots, M$). Minimizing over $\varepsilon^{(k)}_m(e^{j\omega}) = \hat{\Phi}^{(k)}_{u_m z_1}(e^{j\omega}) - \Phi_{u_m z_1}(e^{j\omega})$ in *least squares* (LS) sense, an unbiased estimate for $h_m(e^{j\omega})$ and $l_m(e^{j\omega})$ is obtained.

5. EXPERIMENTAL STUDY

The proposed algorithm was tested in a simulated room environment. The desired and competing speech signals were drawn from TIMIT database, while a recorded fan-noise was used as the stationary noise. All three signals were filtered by simulated room impulse responses, resulting in directional signals, which are received by $M = 5$ microphones. Allen and Berkley's *image method* was used to simulate the ATFs with reverberation time, $T_{60} = 40\,ms$. The length of the filters in the MTFBF, the blocking matrix, and the interference cancellers are set to 250, 250 and 500 taps, respectively. Segments of 1024 samples were used to implement the overlap and save procedure. The sampling frequency is 8KHz. The average SNR of the desired speech signal on all microphones was 6.4 dB. The average SNR of the competing signal on all microphones was 1 dB.

Preliminary results are depicted in Fig. 2. In the top-left and top-right parts a segment of the desired and competing speech signals are illustrated, respectively. Double talk situation is clearly observed. In the bottom-left part the noisy microphone #1 signal is given, while the enhanced signal, after the algorithm has adapted, is depicted in the bottom-right part.

It is clearly seen that the noise level is reduced. Indeed, the desired signal to noise level is increased up to 23.2 dB. The large amount of noise reduction is due to the use of a directional noise source. The competing speech signal during double talk situation is only attenuated by 3.8 dB compared to the desired speech signal. However, the competing speech can be almost completely eliminated in other periods.Informal hearing evaluation confirms that the perceptual quality of the desired speech signal is retained in the enhanced signal, while the stationary and non-stationary interferences are well suppressed.

6. DISCUSSION

We presented a dual source interference canceller based on the TF-GSC, for removing non-stationary directional interference and stationary interference. The MTFBF and the blocking matrix were modified to handle the dual source case. A new system identification was derived for estimating the blocking matrix terms directly, using double talk situations.

The proposed system may be applied to many interesting problems. One possible application is the BSS problem with convolutive mixtures and additive noise. The two sources can be extracted by exchanging the roles of the desired and competing speech signals.

Another application is joint echo cancellation and noise reduction problem, obtained by replacing the competing speech with an echo signal. Note, however, that in this case the input echo signal is available and should be used to improve the obtained performance. This can be done, by incorporating the input echo signal as another input to the system, in a way similar to [6].

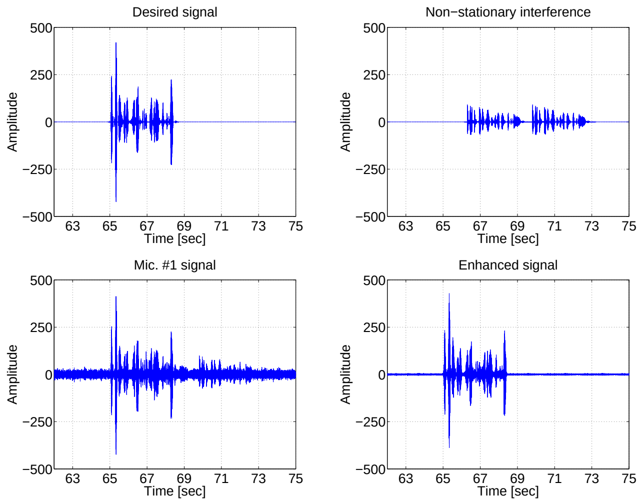

Figure 2: (Top-Left) Desired; (Top-Right) Interference; (Bottom-Left) Mic. #1; (Bottom-Right) Enhanced signal

7. REFERENCES

[1] S. Gannot, D. Burshtein, and E. Weinstein, "Signal Enhancement Using Beamforming and Nonstationarity with Applications to Speech," *IEEE Trans. Signal Processing*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.

[2] L. Parra and C. Spence, "Convolutive Blind Seperation of Non-Stationary Sources," *IEEE transactions on Speech and Audio Processing*, vol. 8, no. 3, pp. 320–327, May 2000.

[3] G. Reuven, S. Gannot, and I. Cohen, "Joint Acoustic Echo Cancellation and Transfer Function GSC in the Frequency Domain," in *Proc. IEEEI*, Herzlia, Israel, Sept. 2004, pp. 412–415.

[4] S. Affes and Y. Grenier, "A Source Subspace Tracking Array of Microphones for Double Talk Situations," in *Proc. ICASSP*, Munich, Germany, Apr. 1996, pp. 269–272.

[5] W. Kellermann, "Strategies for Combining Acoustic Echo Cancellation and Adaptive Beamforming Microphone Arrays," in *Proc. ICASSP*, Munich, Apr. 1997, pp. 219–222.

[6] S. Doclo, M. Moonen, and E. De Clippel, "Combined acoustic echo and noise reduction using gsvd-based optimal filtering," in *Proc. ICASSP*, Istanbul, Turkey, June 2000, pp. 1061–1064.