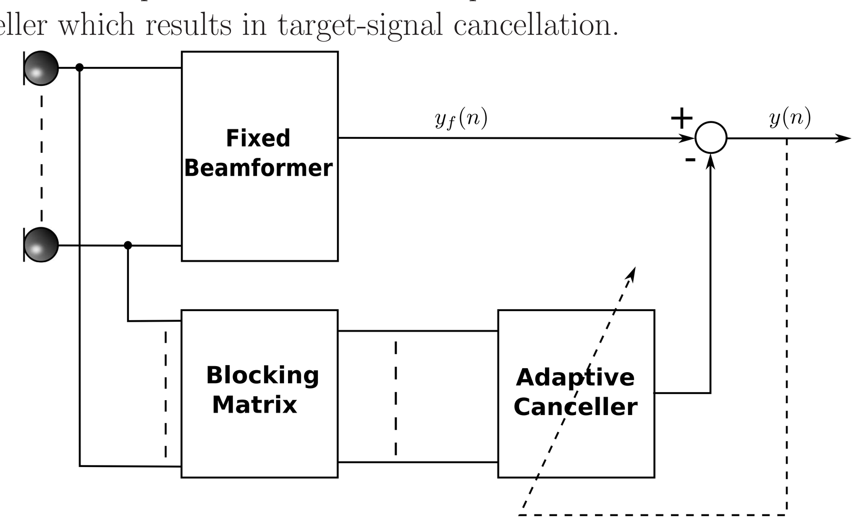
Computationally efficient and robust frequency-domain GSC

Ludovick Lepauloux, Claude Marro
Orange Labs - TECH/OPERA
Lannion, France

Pascal Scalart
ENSSAT - IRISA/CAIRN
Lannion, France

Introduction & Motivation

This work aims at solve the target-signal cancellation problem of the generalized sidelobe canceller (GSC) [1]. In reverberant environments, a single direction of arrival cannot be determined since the desired signal and its reflexion impinge on the array from several directions. Thus, complete rejection of the target-signal is almost impossible in the blocking matrix and a considerable portion of the desired speech will leak to the interference canceller which results in target-signal cancellation.



Generalized Sidelobe Canceller (GSC)

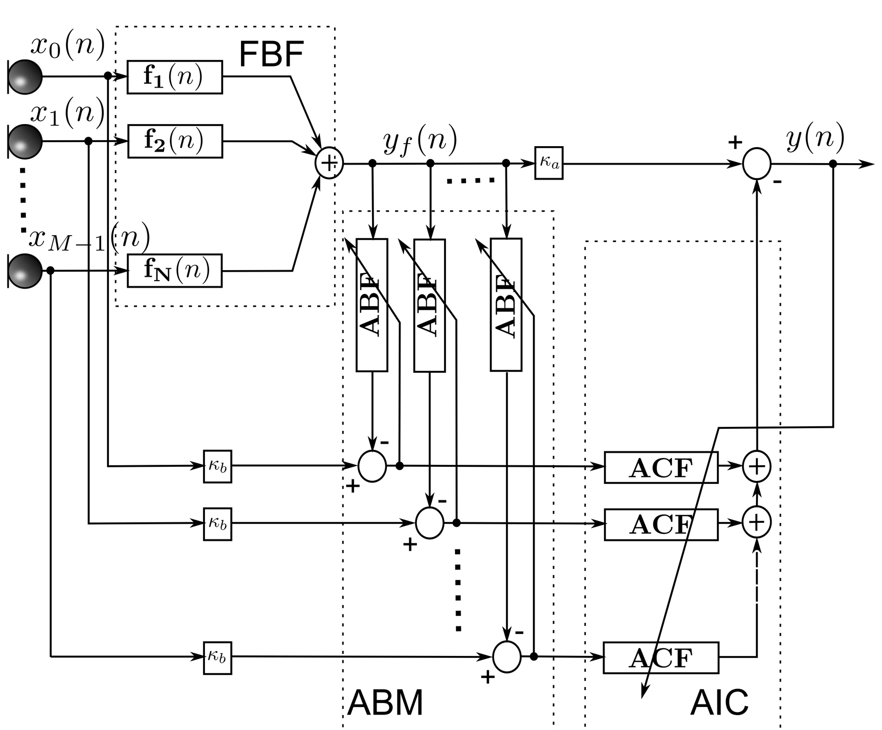
Our goal is to propose a novel adaptive beamforming algorithm with low signal distortion capability and high computational efficiency.

Method

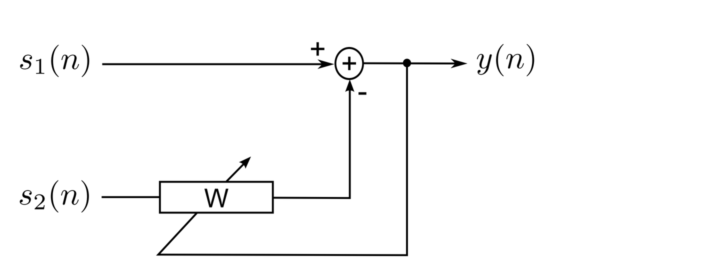
We propose to take benefit of the ability of the crosstalk-resistant adaptive noise canceller (CTRANC) [2] to deal with crosstalk problem which in our case is the same as the signal leakage problem in the GSC. More precisely, we use the CTRANC approach in place of the adaptive interference canceller (AIC)

Our work splits in three main contributions:

1. Define a **new adaptive algorithm in the frequency-domain for the recursive structure** (CTRANC)
2. Derive **new nonparametric variable step-sizes** for each adaptive filter that are easy to control
3. Build the **new computationally efficient recursive GSC** by using the optimal implementation in the frequency-domain [3] of the robust time-domain GSC (RGSC) defined by Hoshuyama et al. in [4]

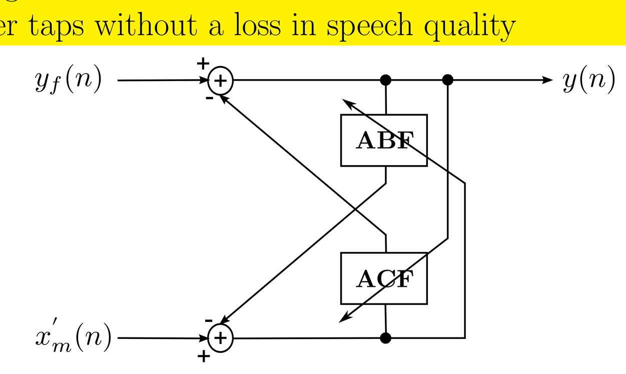


Robust time-domain GSC



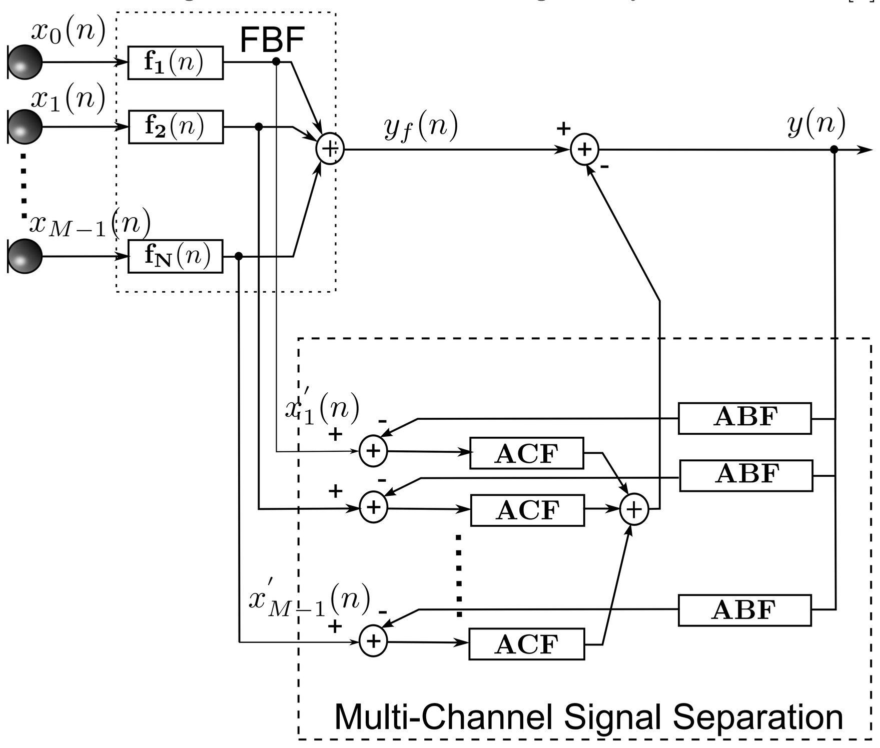
Basic cell constituing the RGSC

1. Replacing the basic feedforward structure by the CTRANC structure:
   - Control of signal leakage leading to improve noise reduction
   - No vocal activity detector
2. Connecting the ABF and the ACF in feedback enables to reduce the number of filter taps without a loss in speech quality



Crosstalk-resistant adaptive noise canceller (CTRANC)

The closest existing work in the literature is given by Choi et al. in [5].



Robust time-domain GSC having its ABF and ACF connected in feedback

Frequency-domain CTRANC and robust GSC

Frequency-domain CTRANC (FCTRANC)

In [6], recursive gradient formula in CTRANC algorithm becomes a convolution operation without a loss in performance.

1. Reduce drastically the memory cost and algorithmic complexity
2. Enable direct frequency-domain implementation

Specifities

1. Convolution and correlation of 3 vectors $\Longrightarrow$ in order to avoid time aliasing, we must use a **block size of $N = 2L$ samples with L the length of the adaptive filters rather than $N = L$ as done classicaly** [7]
2. **The gradient quantities are limited to N correct samples, which corresponds to a truncated version of the theoritical ones**
   - **Potential instabilities** may appear with speech signals [6]
   - **The more significant values are contained in this first N samples**

FCTRANC ALGORITHM BASED ON OVERLAP-SAVE SECTIONING

INITIALIZATION

$$\mathbf{W^1}(0) = [0,\dots,0]^T \qquad \mathbf{GradE_1old}(0) = [0,\dots,0]^T \qquad \mathbf{e_1}(0) = [\mathbf{0_N}]^T$$
$$\mathbf{W^2}(0) = [0,\dots,0]^T \qquad \mathbf{GradE_2old}(0) = [0,\dots,0]^T$$
$$\mathbf{G}(0) = [0,\dots,0]^T \qquad Pr_m(0) = Pf_m(0) = \delta_m \qquad m = 0,\dots,2N-1$$

MATRIX DEFINITIONS:

$$\mathbf{g} = \begin{bmatrix} \mathbf{I_N} & \mathbf{0_N} \\ \mathbf{0_N} & \mathbf{0_N} \end{bmatrix} \qquad \mathbf{k} = [\mathbf{0_N I_N}]\text{; sectioning constraints}$$

$$\mathbf{F} = 2N \times 2N \text{ DFT matrix}$$

FOR EACH NEW BLOCK OF N INPUT SAMPLES:

Signal from the ABM output: $Ob(k)$

$$\mathbf{X_r}(k) = diag\{\mathbf{F}\,[e_1(kN-N),\dots,e_1(kN-1),e_1(kN),\dots,e_1(kN+N-1)]^T\}$$

Convolution
$$\mathbf{Y_2}(k) = \mathbf{X_r}(k)\mathbf{W^2}(k)$$
OLS
$$\mathbf{y_2}(k) = \mathbf{kF^{-1}Y_2}(k)$$
Error calculus
$$\mathbf{ob}(k) = \mathbf{kF^{-1}Ob}(k)$$
$$\mathbf{e_2}(k) = \mathbf{ob}(k) - \mathbf{y_2}(k)$$
$$\mathbf{E_2}(k) = \mathbf{Fk}^T\mathbf{e_2}(k)$$
Step size
$$Pr_m(k) = \lambda\, Pr_m(k-1) + (1-\lambda)\,|X_{r_m}(k)|^2 \qquad m = 0,\dots,2N-1$$
$$\mu_\mathbf{r}(k) = diag\{Pr_0^{-1}(k),\dots,Pr_{2N-1}^{-1}(k)\}$$
Stochastic gradient
$$\mathbf{G}(k) = \mathbf{W^1}(k)\mathbf{W^2}(k)$$
$$\mathbf{W^2}(k{+}1) = \mathbf{W^2}(k) + 2\mu\,\mathbf{FgF^{-1}}\mu_\mathbf{r}(k)\mathbf{E_2}(k)\left[\mathbf{X_r}^H(k) - (\mathbf{G}(k)\mathbf{GradE_2old}(k))^H\right]$$
$$\mathbf{GradE_2old}(k) = \mathbf{FgF^{-1}}\mu_\mathbf{r}(k)\mathbf{E_2}(k)\left[\mathbf{X_r}^H(k) - (\mathbf{G}(k)\mathbf{GradE_2old}(k))^H\right]$$

Signal from the Fixed Beamformer output: $y_f(n)$

$$\mathbf{X_f}(k) = diag\{\mathbf{F}\,[e_2(kN-N),\dots,e_2(kN-1),e_2(kN),\dots,e_2(kN+N-1)]^T\}$$

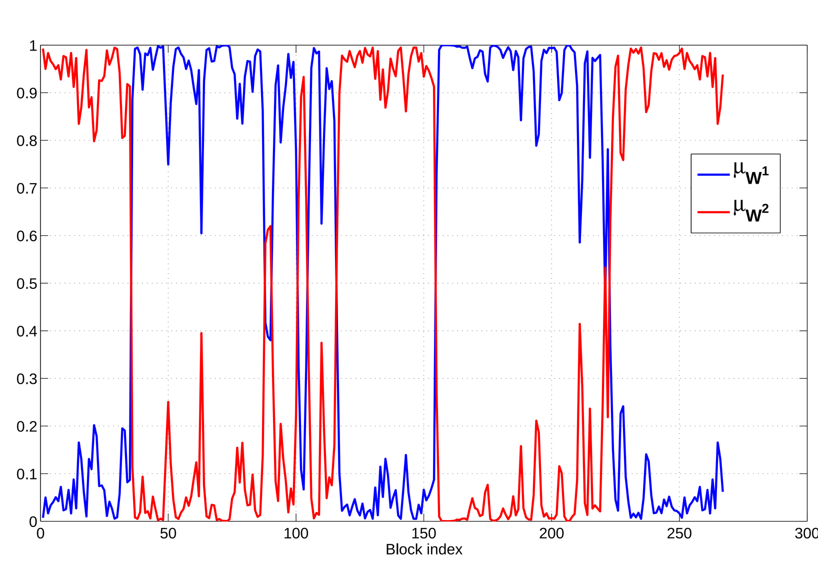
Convolution
$$\mathbf{Y_1}(k) = \mathbf{X_f}(k)\mathbf{W^1}(k)$$
OLS
$$\mathbf{y_1}(k) = \mathbf{kF^{-1}Y_1}(k)$$
Error calculus
$$\mathbf{e_1}(k) = \mathbf{y_f}(k) - \mathbf{y_1}(k)$$
$$\mathbf{E_1}(k) = \mathbf{Fk}^T\mathbf{e_1}(k)$$
Step size
$$Pf_m(k) = \lambda\, Pf_m(k-1) + (1-\lambda)\,|X_{f_m}(k)|^2 \qquad m = 0,\dots,2N-1$$
$$\mu_\mathbf{f}(k) = diag\{Pf_0^{-1}(k),\dots,Pf_{2N-1}^{-1}(k)\}$$
Stochastic gradient
$$\mathbf{W^1}(k{+}1) = \mathbf{W^1}(k) + 2\mu\,\mathbf{FgF^{-1}}\mu_\mathbf{f}(k)\mathbf{E_1}(k)\left[\mathbf{X_f}^H(k) - (\mathbf{G}(k)\mathbf{GradE_1old}(k))^H\right]$$
$$\mathbf{GradE_1old}(k) = \mathbf{FgF^{-1}}\mu_\mathbf{f}(k)\mathbf{E_1}(k)\left[\mathbf{X_f}^H(k) - (\mathbf{G}(k)\mathbf{GradE_1old}(k))^H\right]$$

Step-size control

Due to the non-stationarities of the speech signal the constant step-size 2μ are replaced by two **nonparametric adaptive step-sizes**. These are obtained thanks to the **two-step noise reduction technique** [8].

$$\mu_{\mathbf{W^1}}(k) = \frac{1}{1 + SNR_{prio}(k)}$$
$$\mu_{\mathbf{W^2}}(k) = \frac{SNR_{prio}(k)}{1 + SNR_{prio}(k)}$$

where SNR_{prio} stands for the *a priori* signal-to-noise ratio (SNR) measured on $y_f(n)$.

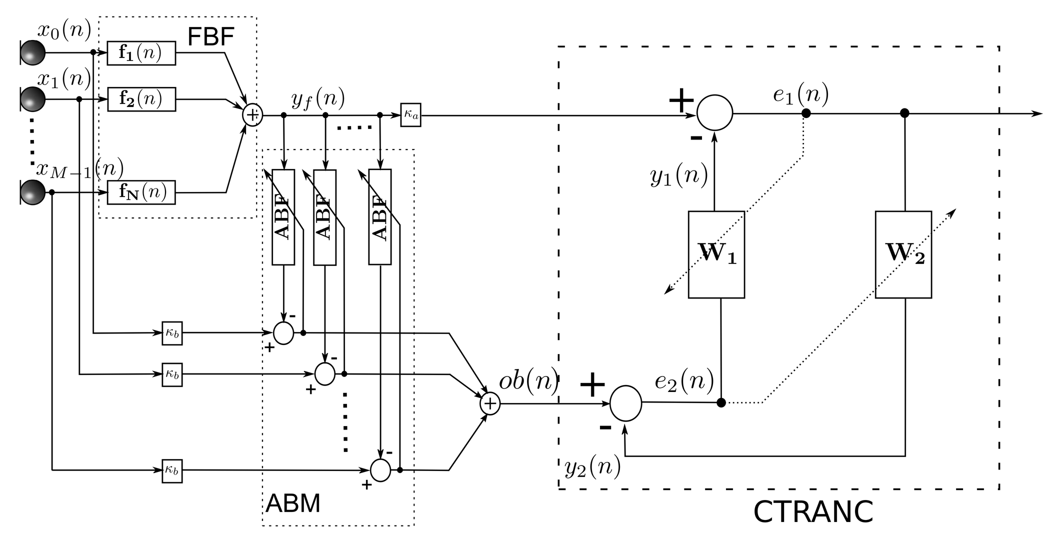


Behavior of the proposed adaptive step-sizes for the frequency bin $k = 1kHz$

- Each step-size evolves in the interval [0,1]
- They are complementary in each frequency bin

Proposed robust GSC

The complete algorithm for the proposed robust GSC is obtained by connecting directly the CTRANC algorithm to the fixed beamformer and ABM outputs defined in [3].

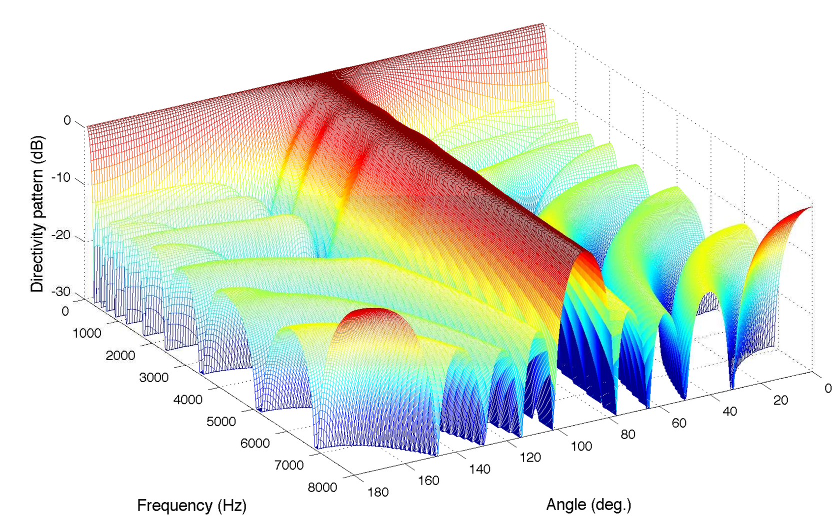


Proposed robust GSC

Experimental results

Array implementation:

Directivity-controlled array made of eleven cardioid microphones clustered in four subarrays is used. The talker location is assumed to be fixed.



Directivity-controlled array made of eleven cardioid microphones

Experimental setup:

- Real videoconferencing context
- Reverberation time varies from 650 ms at 125 kHz to 250 ms at 4 kHz
- Four english sentences (two male voices and two female voices) with a sample rate of 16 kHz in office noise
- Ideal vocal activity detector is used for the estimation of the noise power spectral density

Objective criteria:

- Noise Reduction (**NR**): noise only period
- SNR Gain (**G**): during speech activity
- Cepstral Distance (**CD**): between the direct sound at the central microphone (6) and the output of the system

	NR [dB]				G [dB]			
Herbordt	16.2	13.9	11.9	10.8	**11.8**	**10.8**	**9.5**	**8.6**
Proposed	14.6	12.1	9.8	8.5	9.1	8	6.5	5.6
Choi	**16.7**	13.4	12.3	13.2	6.8	5.6	5.4	5.8
Choi (modified version)	14.7	**15**	**14.5**	**13.5**	8.7	9	8.7	8.1
$\mathbf{SNR_{in}}$ [dB]	0	5	10	15	0	5	10	15

	CD			
Herbordt	0.54	0.53	0.5	0.47
Proposed	0.44	0.42	0.42	0.41
Choi	0.41	0.38	0.36	**0.36**
Choi (modified version)	**0.38**	**0.38**	**0.36**	0.39
$\mathbf{SNR_{in}}$ [dB]	0	5	10	15

Conclusions:

1. The proposed robust GSC offers a **drastic complexity gain** and a **lower signal distortion in comparison with the direct structure given by Herbordt**, but at the expense of a **slight loss in noise reduction**
2. **Recursive implementations offer the higher speech quality**
3. Choi (modified version) outperforms the other methods, especially the one proposed by Choi $\Longrightarrow$ **Our algorithm takes into account the cross-coupling effect of the CTRANC approach**

References

[1] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," in *IEEE Trans. on Antennas and Propagation*, 1982.

[2] G. Mirchandani, R. L. Zinser, Jr., and J. B. Evans, "A new adaptive noise cancellation scheme in the presence of crosstalk," *IEEE Trans. on circuits and systems II: Analog and Digital Signal Processing*, vol. 39, pp. 681–694, 1992.

[3] W. Herbordt and W. Kellermann, "Efficient frequency-domain realization of robust generalized sidelobe cancellers," in *IEEE fourth Workshop on Multimedia Signal Processing*, 2001.

[4] O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," in *IEEE Trans. on Signal Processing*, 1999.

[5] C. Choi, D. Kong, J. Kim, and S. Bang, "Speech enhancement and recognition using circular microphone array for service robots," in *IROS*, 2003.

[6] L. Lepauloux, P. Scalart, and C. Marro, "An efficient low-complexity algorithm for crosstalk-resistant adaptive noise canceller," in *EUSIPCO*, 2009.

[7] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Processing Magazine*, vol. 9, pp. 14–37, 1992.

[8] C. Plapous, C. Marro, L. Mauuary, and P. Scalart, "A two-step noise reduction technique," in *ICASSP*, 2004.