Performance analysis of MVDR beamformer in WASN with sampling rate offsets and blind synchronization

Conference Paper · September 2015

DOI: 10.1109/EUSIPCO.2015.7362382

3 authors:

Dani Cherkassky
Bar Ilan University
10 PUBLICATIONS **14** CITATIONS
SEE PROFILE

Shmulik Markovich-Golan
Bar Ilan University
20 PUBLICATIONS **184** CITATIONS
SEE PROFILE

Sharon Gannot
Bar Ilan University
172 PUBLICATIONS **2,173** CITATIONS
SEE PROFILE

PERFORMANCE ANALYSIS OF MVDR BEAMFORMER IN WASN WITH SAMPLING RATE OFFSETS AND BLIND SYNCHRONIZATION

Dani Cherkassky, Shmulik Markovich-Golan and Sharon Gannot

Faculty of Engineering
Bar-Ilan University
Ramat-Gan, 5290002, Israel

dani.cherkassky@gmail.com; shmuel.markovich@biu.ac.il; sharon.gannot@biu.ac.il

ABSTRACT

In wireless acoustic sensor networks (WASNs), sampling rate offsets (SROs) between nodes are inevitable, and recognized as one of the challenges that have to be resolved for a coherent array processing. A simplified free-space propagation is considered with a single desired source impinging a WASNs from the far-field and contaminated by a diffuse noise. In this paper, we analyze the theoretical performance of a fixed superdirective beamformer (SDBF) in presence of SROs. The SDBF performance loss due to SROs is manifested as a distortion of the nominal beampattern and an excess noise power at the output of the beamformer. We also propose an iterative algorithm for SROs estimation.The theoretical results are validated by simulation.

***Index Terms*—** Blind synchronization, Wireless acoustic sensor network, Sampling rate offset

1. INTRODUCTION

The use of wireless acoustic sensor network (WASN) as a speech processing tool has recently attracted a significant research attention. Along with the appealing advantages offered by WASNs, some new challenges arise. One of the challenges is the synchronization between WASN nodes. Contrary to a centralized microphones array, where all signals are sampled with the same clock, the sampling process in each WASN node relies on its local clock source, thus, sampling rate offsets (SROs) are inevitable.

Clock synchronization in distributed sensors network has been addressed in the literature, in a wider context than speech processing - e.g. in [1], [2]. This important topic was also considered in general speech/audio processing applications, such as echo cancelation [3], and blind source separation [4]. In WASN, the synchronization techniques can be classified in two groups: the *time stamps* approach which utilizes the communication links between the sensors to distribute synchronization data in the network, and a *blind* approach which only utilizes the acoustic signals. Early works using the *time stamps* approach are [5] and [2]. Recently, a comprehensive study was presented in [6], where synchronization is carried out using combined hardware and software methods. The *blind* approach was also a subject for a considerable amount of research [7], [8], [9]. The general idea is to model the distortion imposed by the SRO on the audio signals and to estimate the SROs in the WASN with respect to (w.r.t.) an acoustical signal from a reference node.

In the current contribution, we adopt the *blind* approach. The SRO effect is modeled as a time-varying delay between the signals. Using this approximation we theoretically analyze the SDBF beampattern, and the excess noise power at the output of the SDBF in presence of SRO. Furthermore, we propose an iterative algorithm for SRO estimation. The algorithm is based on maximizing the coherence between the WASN signals, in the short-time Fourier transform (STFT) domain.

The rest of the paper is organized as follows. In Sec. 2 the problem is formulated. In Sec. 3 SDBF performance degradation due to SRO is analyzed. In Sec. 4 we describe the SROs estimation algorithm. The performance of the proposed methods is evaluated in Sec. 5. We conclude this paper by a short discussion in Sec. 6.

2. PROBLEM FORMULATION

Consider a WASN comprising M microphones, aiming at enhancing a desired speech signal in the presence of a spherically isotropic noise field, also known as diffuse noise field. In the STFT domain, the speech signal is denoted $s(l,k)$, the speech steering vector is denoted $\mathbf{g}(k)$, and the noise at the mth microphone is denoted $v_{\mathrm{m}}(l,k)$, where l is the frame index, and $k = 0,\ldots,K-1$ is the frequency index. The STFT analysis window length is denoted L. The speech enhancement is accomplished by applying the minimum variance distortionless response (MVDR) beamformer. In the sequel, the term *nominal* will correspond to values used for designing the beamformer. The nominal array signal is given by:

$$\mathbf{z}^{\mathrm{n}}(l,k) = \mathbf{g}(k)s(l,k) + \mathbf{v}(l,k). \tag{1}$$

For the sake of simplicity, we assume in the sequel that each node comprises a single microphone. Denote the sampling rate at the mth microphone as $f_{s,m}$. Without loss of generality, the sampling rate of the mth node is defined in terms of the sampling rate of the first (reference) node, f_s, as $f_{s,m} = f_s/a_m$ where $a_1 = 1$. The identity between the SRO phenomenon and time-scaling was introduced in [9]. Using this identity, we can formulate the actual array signal as:

$$\mathbf{z}^{\mathrm{u}}(l,k) = \begin{bmatrix} \tilde{s}_1(l,k) & 0 & \cdots \\ 0 & \ddots & 0 \\ \vdots & 0 & \tilde{s}_M(l,k) \end{bmatrix} \begin{bmatrix} \tilde{g}_1(k) \\ \vdots \\ \tilde{g}_M(k) \end{bmatrix} + \begin{bmatrix} \tilde{v}_1(l,k) \\ \vdots \\ \tilde{v}_M(l,k) \end{bmatrix}, \tag{2}$$

where $\tilde{s}_{\mathrm{m}}(l,k)$ and $\tilde{v}_{\mathrm{m}}(l,k)$ are the STFTs of the time-scaled, continuous-time, speech and the noise signals, $s(a_m t)$ and $v_{\mathrm{m}}(a_m t)$, respectively with t denoting the continuous-time axis. Similarly, $\tilde{g}_{\mathrm{m}}(k)$ is the discrete Fourier transform of the time-scaled, sampled, impulse response between the speech and the mth sensor $g_{\mathrm{m}}(a_m t)$.

For the sake of simplicity, we are considering a free-field scenario and a linear microphone array consisting of M microphones with d_m being the distance between the mth microphone and the reference (first) microphone. Although we assume a linear array in this paper, all the results can be readily extended to an arbitrary three-dimensional WASN configuration. The speech signal is impinging on the array from a far-field with a direction of arival (DOA) denoted by θ. Accordingly, the MVDR which is a SDBF in our case, is solely defined by the geometrical properties of the setup:

$$\mathbf{w}(k) = \frac{\mathbf{\Gamma}_{\mathbf{vv}}^{-1}(k)\mathbf{g}(k)}{\mathbf{g}^H(k)\Gamma_{\mathbf{vv}}^{-1}(k)\mathbf{g}(k)}, \tag{3}$$

where $\mathbf{\Gamma}_{\mathbf{vv}}(k)$ is the spatial coherence matrix of the diffuse noise with $[\mathbf{\Gamma}_{\mathbf{vv}}(k)]_{i,j} = \mathrm{sinc}\left(\frac{2\pi(d_i - d_j)}{\lambda_k}\right)$, $\lambda_k = \frac{c}{f_s}\frac{K}{k}$ is the wavelength corresponding to the kth frequency index, c is the sound velocity in the medium, and the steering vector $\mathbf{g}(k)$ is given by:

$$\mathbf{g}(k) = \left[e^{-j2\pi\frac{d_1}{\lambda_k}\cos(\theta)}, \cdots, e^{-j2\pi\frac{d_M}{\lambda_k}\cos(\theta)}\right]^T. \tag{4}$$

In the scope of this work we examine the performance of $\mathbf{w}(k)$ when applied to the unsynchronized signal $\mathbf{z}^{\mathrm{u}}(l,k)$, and propose a method for estimating the SROs $\{a_m\}_{m=2}^{M}$.

3. SDBF PERFORMANCE ANALYSIS WITH SRO

We turn now to the derivation of a simplified expression for $\mathbf{z}^{\mathrm{u}}(l,k)$. We consider the mth microphone signal as a function of the continuous-time t_m, with time axis t_m related to the time axis of the first microphone signal by $t_m = a_m t_1$, and $a_m = 1 + \epsilon_m$ is the respective SRO. The continuous-time t_m can be defined in terms of t_1, as proposed in [8]:

$$\begin{aligned} t_m &= (1+\epsilon_m)t_1 = (1+\epsilon_m)(t_1 - T_l) + (1+\epsilon_m)T_l = \\ &= t_1 - T_l + \epsilon_m(t_1 - T_l) + \epsilon_m T_l + T_l, \end{aligned} \tag{5}$$

where T_l is the center of the lth frame at the first microphone. Considering the lth frame, t_1 is within the range of $T_l - \frac{L}{2f_s} \leq t_1 \leq T_l + \frac{L}{2f_s}$, and hence $|\epsilon_m|(t_1 - T_l) \leq |\epsilon_m|\frac{L}{2f_s}$. Assuming both ϵ_m and L are sufficiently small, the term $\epsilon_m(t_1 - T_l)$ can be neglected, resulting in:

$$t_m - T_l \approx t_1 - T_l + \epsilon_m T_l. \tag{6}$$

Based on (6), and the properties of the STFT we can approximate the SRO effect on the speech and the noise signals in the STFT domain as follows:

$$\tilde{s}_{\mathrm{m}}(l,k) \approx s(l,k)e^{j2\pi f_s \frac{k}{K}\epsilon_m T_l} = s(l,k)e^{j\pi k\epsilon_m l}, \tag{7a}$$

$$\tilde{v}_{\mathrm{m}}(l,k) \approx v_{\mathrm{m}}(l,k)e^{j2\pi f_s \frac{k}{K}\epsilon_m T_l} = v_{\mathrm{m}}(l,k)e^{j\pi k\epsilon_m l}, \tag{7b}$$

while the rightmost term in (7a) and (7b) are obtained, without loss of generality, for analysis window length of $L = K$, and 50% overlap between successive frames. In this case, we can substitute: $T_l = \frac{L}{2f_s}l$.

Consider $g_{\mathrm{m}}(k) = e^{-j2\pi\frac{d_m}{\lambda_k}\cos(\theta)}$, the discrete Fourier transform of the sampled, nominal, impulse response between the speech source and the mth sensor $g_{\mathrm{m}}(t)$. Due to SRO, the time axis of the impulse response is scaled by a_m. When applying the discrete Fourier transform to the time-scaled, and sampled impulse response, the kth frequency index corresponds to a continuous-time signal with a wavelength of $a_m\lambda_k$, instead of λ_k in the nominal case. An approximate expression for $\tilde{g}_{\mathrm{m}}(k)$ results in by replacing the term $1/a_m$ with its first-order Taylor series approximation $1 - \epsilon_m$:

$$\begin{aligned} \tilde{g}_{\mathrm{m}}(k) &= e^{-j2\pi\frac{d_m}{a_m\lambda_k}\cos(\theta)} \approx \\ &\approx e^{-j2\pi\frac{d_m}{\lambda_k}\cos(\theta)} \cdot e^{j2\pi\frac{d_m}{\lambda_k}\cos(\theta)\epsilon_m}. \end{aligned} \tag{8}$$

A simplified expression for $\mathbf{z}^{\mathrm{u}}(l,k)$ is obtained by substituting (7a), (7b), and (8) in (2):

$$\mathbf{z}^{\mathrm{u}}(l,k) \approx \mathbf{E}_{\mathrm{s}}(l,k)\mathbf{g}(k,\theta)s(l,k) + \mathbf{E}(l,k)\mathbf{v}(l,k) \tag{9}$$

where the SRO effect is modeled by the diagonal matrices $\mathbf{E}_{\mathrm{s}}(l,k)$ and $\mathbf{E}(l,k)$:

$$\mathbf{E}_{\mathrm{s}}(l,k) = \begin{bmatrix} e^{j\pi\left(\frac{2d_1}{\lambda_k}\cos(\theta)+kl\right)\epsilon_1} & 0 & \cdots \\ 0 & \ddots & 0 \\ \vdots & 0 & e^{j\pi\left(\frac{2d_M}{\lambda_k}\cos(\theta)+kl\right)\epsilon_M} \end{bmatrix}, \tag{10}$$

$$\mathbf{E}(l,k) = \begin{bmatrix} e^{j\pi kl\epsilon_1} & 0 & \cdots \\ 0 & \ddots & 0 \\ \vdots & 0 & e^{j\pi kl\epsilon_M} \end{bmatrix}. \tag{11}$$

Consider the effect of the SRO on the beampattern of the SDBF applied to $\mathbf{z}^{\mathrm{u}}(l,k)$. The nominal beampattern of the SDBF is defined by $B^{\mathrm{n}}(k,\theta) = \mathbf{w}^H(k)\mathbf{g}(k,\theta)$. Accordingly, by comparing (9) to (1), the beampattern of the nominal beamformer $\mathbf{w}(k)$ applied to the unsynchronized signals is given by:

$$B(l,k,\theta) = \mathbf{w}^H(k)E_{\mathrm{s}}(l,k)\mathbf{g}(k,\theta). \tag{12}$$

Note, that the beampattern in the unsynchronized case is time-dependent, due to the drifting delay. The excess noise power at the output of the unsynchronized beamformer is defined as:

$$R \triangleq \frac{E\left\{\|\mathbf{v}^{\mathrm{sro}}\|^2\right\}}{E\left\{\|\mathbf{v}^{\mathrm{n}}\|^2\right\}} = \frac{\mathbf{w}^H(k)\mathbf{E}(l,k)\mathbf{\Gamma}_{\mathrm{vv}}(k)\mathbf{E}^H(l,k)\mathbf{w}(k)}{\mathbf{w}^H(k)\mathbf{\Gamma}_{\mathrm{vv}}(k)\mathbf{w}(k)} = \\ = \gamma(l,k)\cdot\frac{1}{\mathbf{g}^H(k)\mathbf{\Gamma}_{\mathrm{vv}}^{-1}(k)\mathbf{g}(k)}, \tag{13}$$

where:

$$\gamma(l,k) = \mathbf{g}^H(k)\mathbf{\Gamma}_{\mathrm{vv}}^{-1}(k)\mathbf{E}(l,k)\mathbf{\Gamma}_{\mathrm{vv}}(k)\mathbf{E}^H(l,k)\mathbf{\Gamma}_{\mathrm{vv}}^{-1}(k)\mathbf{g}(k). \tag{14}$$

In conclusion, the beampattern of $\mathbf{w}(k)$ applied to $\mathbf{z}^{\mathrm{u}}(l,k)$, and the excess noise power due to SROs are given by equations ( ) and ( ), respectively.

4. SRO ESTIMATION AND COMPENSATION

In this section we propose an algorithm for estimating $\{a_m\}_{m=2}^M$. The algorithm is based on maximizing the coherence between the reference node signal and the mth node signal in the STFT domain. The effectiveness of coherence-based algorithm was shown in [7], [8], and [9]. Here we present an iterative algorithm based on the approximate SRO model (9). We start by noticing that in practical network configurations $\frac{2d_m}{\lambda_k}\cos(\theta) + kl \approx kl$ for $l \gg 1$. Thus, for $l \gg 1$ we observe that $\mathbf{E}_{\mathrm{s}}(l,k) \approx \mathbf{E}(l,k)$ and the array signal can be further approximated by:

$$\mathbf{z}^{\mathrm{u}}(l,k) \approx \mathbf{E}(l,k)\left(\mathbf{g}(k,\theta)s(l,k) + \mathbf{v}(l,k)\right). \tag{15}$$

From (15) it is clear that the performance degradation due to SRO can be compensated by multiplying the array signal by $\mathbf{E}^{-1}(l,k)$. Since $\mathbf{E}^{-1}(l,k)$ is diagonal the SROs w.r.t. the reference (first) node can be compensated at each node independently. The compensation at the mth node is carried out by multiplying its signal by $[\mathbf{E}^{-1}(l,k)]_{mm} = e^{-j\pi kl\epsilon_m}$. In practice, the parameters $\{\epsilon_m\}_{m=2}^M$ are unknown, and should be estimated from the data. Since the coherence between synchronized signals is higher than the respective coherence of the non-synchronized signals, even in a reverberant and multiple speakers scenario [9], we can obtain an estimate of the SRO by maximizing the coherence. Accordingly, an estimate for ϵ_m can be derived by solving the following optimization problem:

$$\hat{\epsilon}_m = \underset{\epsilon'_m}{\mathrm{argmax}}\left\{|J_m(k,\epsilon'_m)|^2\right\}, \tag{16}$$

$$J_m(k,\epsilon'_m) \triangleq \frac{1}{L_J}\sum_{l=0}^{L_J-1}(z_1^{\mathrm{u}}(l,k))^*\cdot z_m^{\mathrm{u}}(l,k)e^{-j\pi kl\epsilon'_m}, \tag{17}$$

where J_m is the sample covariance between the reference signal and the signal from the mth node with SRO compensation of ϵ'_m, and L_J is the number of frames used for calculating J_m. We are not familiar with a closed-form solution to (16). However, the optimization of the cost function w.r.t. ϵ'_m can be carried out using the *gradient acescent* method:

$$\hat{\epsilon}_m(i+1) = \hat{\epsilon}_m(i) + \beta\cdot\nabla_{\hat{\epsilon}_m}|J_m(k,\hat{\epsilon}_m(i))|^2, \tag{18}$$

with β being the stepsize, and i the iteration index. The gradient term is obtained by calculating the derivative of J_m w.r.t. ϵ'_m and straightforward algebra:

$$\nabla_{\epsilon'_m}|J_m(k,\epsilon'_m)|^2 = \frac{\partial}{\partial\epsilon'_m}|J_m(k,\epsilon'_m)|^2 = \\ = \frac{2\pi k}{L_J}\mathrm{Im}\left\{J_m^*(k,\epsilon'_m)\sum_{l=0}^{L_J-1} l\cdot(z_1^{\mathrm{u}}(l,k))^*\cdot z_m^{\mathrm{u}}(l,k)e^{-j\pi kl\epsilon'_m}\right\}. \tag{19}$$

From the estimates of $\{\epsilon_m\}_{m=2}^M$, an estimate of $\mathbf{E}^{-1}$ is deduced by $\widehat{[\mathbf{E}^{-1}(l,k)]}_{mm} = e^{-j\pi kl\hat{\epsilon}_m}$. The SRO can be compensated by multiplying the array signal with $\widehat{E^{-1}(l,k)}$. In general, estimation is never error-free, hence, even after compensating for the SRO with the proposed technique, (12) and (13) can be used to analyze the performance degradation due to the SRO estimation error, in this case $\{\epsilon_m\}_{m=2}^M$ will represent the uncompensated residual SRO rather than the SRO itself.

5. EXPERIMENTAL STUDY

In this section we verify the methods proposed in Sec. 3 and Sec. 4. For that purpose, a simulative benchmark has been designed. Two simulative studies were carried out. The first is aiming at verifying the SDBF beampattern, and the excess noise power models derived in (12), and (13), respectively. The second study is aiming at analyzing the performance of the SRO estimation algorithm (18).

5.1. SRO effect on SDBF

In this study, a uniform linear array with $M = 4$ microphones spaced by 11 cm was used. A speech signal impinging on the array from the far-field with DOA equal to $\theta = 60^0$ is simulated in a noise-free environment. The unsynchronized signals were generated by re-sampling the synchronized signals with randomly selected SROs, denoted SRO_m,

for $m = 2, \ldots, M$. The average SRO is denoted $\overline{\text{SRO}} = \frac{1}{M-1}\sum_{m=2}^{M} \text{SRO}_m$. The array gain towards an unsynchronized signal arriving from $0^0 \leq \theta \leq 180^0$ was calculated as a power ratio of the signal at the output of the SDBF and the signal of one of the microphones. This empirical gain is denoted by B^{sim}.

A comparison between $B^{\text{sim}}(\theta)$ and $B(\theta)$ as defined by (12) is presented in Fig. 1. Clearly, the analytically derived beampattern is in a very good agreement with the empirical one. The effect of the beampattern variation in time is also exemplified in Fig. 1, by comparing the upper and the lower figures. The nominal beampattern of the SDBF B^{n} and the average (over time) beampattern $\overline{B}$ due to $\overline{\text{SRO}} = 100$ PPM are also depicted in both figures in Fig. 1.

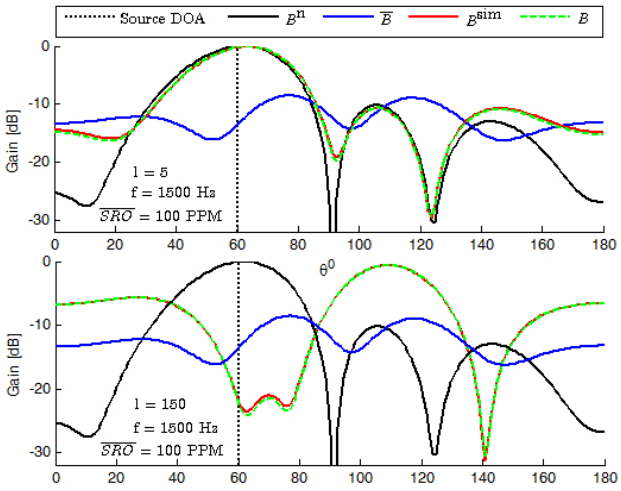

Fig. 1. Empirical vs. analytical beampatterns of SDBF.

The sensitivity of the SDBF to SROs is analyzed in Fig. 2. As seen from the upper figure, the average beampattern due to $\overline{\text{SRO}} = 1$ PPM is reasonably similar to the nominal one. However, the average beampattern due to a modest imperfection of $\overline{\text{SRO}} = 10$ PPM is far from the nominal one, as shown in the bottom plot. Note that, the level of the SRO affect the rate at which the beampattern varies. However, even for the smallest SRO, after enough time has elapsed the beampattern will be very different from the nominal one.

To evaluate the excess noise power model (13), a diffuse noise $\mathbf{v}^{\text{n}}(l,k)$ was generated by the procedure proposed in [10]. SROs with $\overline{\text{SRO}} = 100$ PPM were introduced to the synchronized noise signals, resulting in $\mathbf{v}^{\text{u}}(l,k)$. The empirical excess noise power R^{sim} was calculated as the power ratio between SDBF response to $\mathbf{v}^{\text{n}}(l,k)$ and SDBF response to $\mathbf{v}^{\text{u}}(l,k)$. A comparison between R^{sim} and R as defined by (13) is summarized in Table 1. It is interesting to note that the noise reduction at a high frequency is almost unaffected by the SRO. This is due to the characteristic of the diffuse noise, which is known to become incoherent at higher frequencies [10], and hence its reduction is not affected by the

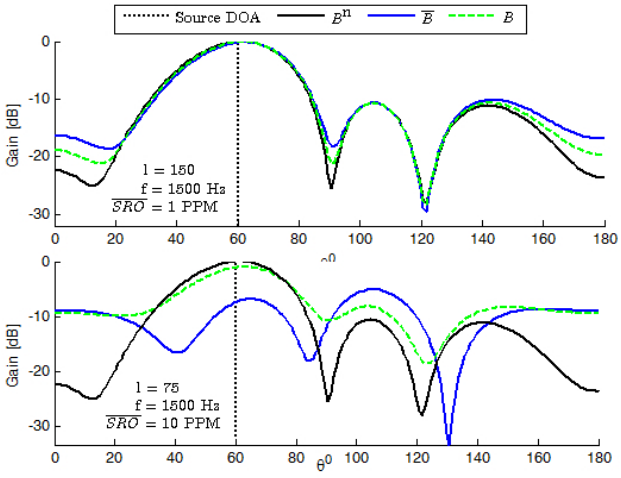

Fig. 2. SDBF beampatterns for various SROs.

SRO. On the contrary, the noise reduction at low frequencies is degrading with time.

f[Hz]	l	R^{sim}[dB]	R[dB]
300	5	2	1
1500	5	0.3	0
300	150	16	18
1500	150	-0.4	0

Table 1. Analytical and empirical excess noise power at the output of the SDBF.

5.2. Iterative estimation of SRO

In this part of the work, we are evaluating the proposed SRO estimation technique (18). The experimental setup is similar to the one described in Sec. 5.1, with an exception of both diffused noise, and speech signal being simultaneously present. The signal to noise ratio (SNR) was set to 10 dB. In the sequel we discuss the SRO estimation for the second microphone ($m = 2$) as an example. Similar procedure can be applied to any microphone $m = 2, \ldots, M$. The SRO was set to 100 PPM. The analysis window is set to $L = 1024$, and $L_J = 150$ frames were used for estimating the SRO.

The performance of the gradient acescent procedure is depicted in Figs. 3 and 4. It is readily observed that the cost function $|J_2(k, \epsilon_2)|^2$ is, in general, non-convex. Hence, convergence of (18) to the global maximum cannot be guaranteed. Indeed, the procedure is trapped in a local maximum due to poor initialization, as shown in Fig. 3.

A successful estimation of ϵ_2 is depicted in Fig. 4. It is easily verified that the cost function is getting smoother when a lower frequency band is considered, which facilitates sufficient estimation performance in this example. It should be

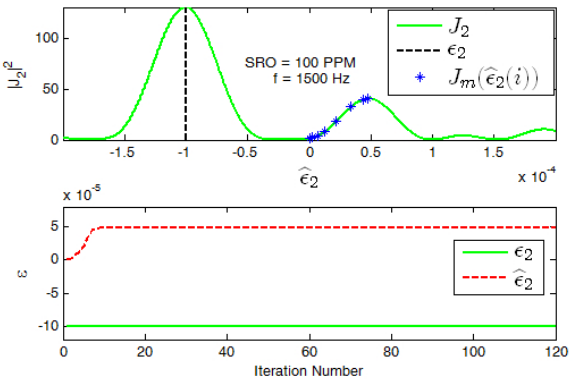

Fig. 3. *Gradient Acescent* trapped in a local maximum due to poorly chosen initial conditions.

noted that the slope of the cost function is getting moderate once a very low frequency band is considered, which also hampers the estimation performance. The above observations dictate that a proper frequency selection methodology is required in order to successfully estimate the SROs using the proposed iterative method. However, such a methodology is beyond the scope of the current contribution and will be a subject for a future study.

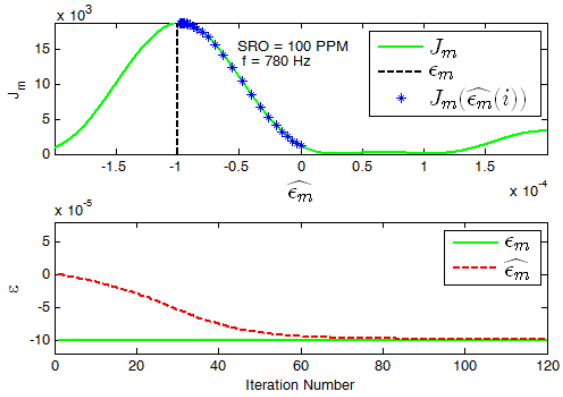

Fig. 4. Cost function $|J_2|^2$ and the *Gradient Acescent* learning curve.

6. SUMMARY

A synchronization between WASN nodes is considered in this work. The SDBF performance loss due to SRO is manifested as a distortion of the nominal beampattern and an excess noise power at the output of the beamformer. An iterative technique for SROs estimation was proposed. The methods and techniques presented were validated by simulations and their limitations were exemplified.

REFERENCES

[1] Yik-Chung Wu, Qasim Chaudhari, and Erchin Serpedin, "Clock synchronization of wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 124–138, 2011.

[2] Luca Schenato and Federico Fiorentin, "Average timesync: A consensus-based protocol for time synchronization in wireless sensor networks," in *Estimation and Control of Networked Systems*, 2009, vol. 1, pp. 30–35.

[3] Matthias Pawig, Gerald Enzner, and Peter Vary, "Adaptive sampling rate correction for acoustic echo control in voice-over-IP," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 189–199, 2010.

[4] Stefan Wehr, Igor Kozintsev, Rainer Lienhart, and Walter Kellermann, "Synchronization of acoustic sensors for distributed ad-hoc audio networks and its use for blind source separation," in *the 6th IEEE International Symposium on Multimedia Software Engineering*, 2004, pp. 18–25.

[5] Raj Thilak Rajan and Alle-Jan van der Veen, "Joint ranging and clock synchronization for a wireless network," in *the 4th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2011, pp. 297–300.

[6] Joerg Schmalenstroeer, Patrick Jebramcik, and Reinhold Haeb-Umbach, "A combined hardware–software approach for acoustic sensor network synchronization," *Signal Processing*, vol. 107, pp. 171–184, 2015.

[7] Shmulik Markovich-Golan, Sharon Gannot, and Israel Cohen, "Blind sampling rate offset estimation and compensation in wireless acoustic sensor networks with application to beamforming," in *International Workshop on Acoustic Signal Processing (IWAENC)*, 2012.

[8] Shigeki Miyabe, Nobutaka Ono, and Shoji Makino, "Blind compensation of inter-channel sampling frequency mismatch with maximum likelihood estimation in STFT domain," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 674–678.

[9] Dani Cherkassky and Sharon Gannot, "Blind synchronization in wireless sensor networks with application to speech enhancement," in *the 14th International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2014, pp. 183–187.

[10] Emanuël A.P. Habets and Sharon Gannot, "Generating sensor signals in isotropic noise fields," *The Journal of the Acoustical Society of America*, vol. 122, no. 6, pp. 3464–3470, 2007.