DISTRIBUTED DELAY AND SUM BEAMFORMER FOR SPEECH ENHANCEMENT IN WIRELESS SENSOR NETWORKS VIA RANDOMIZED GOSSIP

Yuan Zeng and Richard C. Hendriks

Delft University of Technology
Signal and Information Processing Lab
2628 CD Delft, The Netherlands

ABSTRACT

In this paper, we describe a distributed delay and sum beamformer (DDSB) for speech enhancement based on a randomized gossip algorithm. The proposed algorithm operates in a randomly connected wireless sensor network. Without any network topology constraint, the DDSB estimates the desired signal at each node by only exchanging information with its neighbors. Since the DDSB performs only local signal processing, it is robust and scalable for large sensor networks and dynamic environments. We show that the DDSB converges to the optimal estimate of the centralized beamformer. Furthermore, we provide a bound for the worst-case averaging time of the DDSB for the worst connected network. The simulation results validate the theoretical results of the algorithm.

***Index Terms*—** Distributed delay and sum beamformer, randomized gossip, speech enhancement, wireless sensor networks.

1. INTRODUCTION

Speech enhancement algorithms can be used to improve the quality and intelligibility of speech in noisy environments for applications like mobile telephony, hearing aids, human-machine communication systems, etc. While improvements are generally modest for single-microphone algorithms, multi-microphone algorithms can potentially obtain much larger quality and intelligibility improvements by constructing a beamformer. Conventional centralized speech enhancement beamforming algorithms consider generally a relatively small number of microphones and process the data centrally. Advances in micro electro-mechanical systems, enable the use of many low-cost microphones each having their own individual processor in a wireless sensor network (WSN). For a large WSN, the centralized beamforming processing, is neither robust nor scalable, since a single point of failure exists. An alternative is to use distributed algorithms, e.g., [1][2][3], where each node only exchanges data with its neighbors and performs local processing of the data. The distributed algorithms provide robust and scalable solutions for large networks and unreliable communication environments.

In many situations, distributed estimation algorithms are assumed to operate in a network with a specific topology, such as an ideal fully connected topology [1] or a tree topology [2]. These algorithms need higher communication costs, as specialized routes need to be established, and, are not robust for a changing communication environment. Without any specialized network routing requirement, the randomized gossip algorithm [4] is attractive for in-network signal processing. This distributed processing scheme is iterative and uses simple computations. In this paper, we investigate the use of asynchronous randomized gossip within beamforming for speech enhancement. More specifically, as opposed to the traditional centralized delay and sum beamformer (CDSB) we present a distributed delay and sum beamformer (DDSB) which operates in a randomly connected sensor network via gossip processing. In the DDSB, each node estimates the desired signal by using only local information and by performing only local processing, thereby, overcoming the need to transmit data to other then neighboring nodes. In addition, the nodes only need to perform relatively simple operations, putting very low requirements on the node's processor. As the DDSB is asynchronous, needs only local communication and performs local processing, there is no requirement for a special topology and there is no risk of having a single point of failure.

The proposed DDSB is based on an iterative scheme that converges to the optimal solution of the centralized beamformer. Notice that the DDSB is a special case of an MVDR beamformer assuming the noise is uncorrelated across microphones. This is validated for diffuse noise fields and/or when the distance between microphones is sufficiently large. We prove that the DDSB converges asymptotically to the centralized beamformer and derive a bound for the averaging time in the case of the worst connected network. The algorithm is adaptive for real-time speech enhancement in dynamic communication environments without any topological constraints.

This paper is organized as follows. In Sec. 2, the estimation problem is formulated and notation is given. Then, in Sec. 3, we discuss the optimal centralized beamformer. In Sec. 4, we briefly review the randomized gossip algorithm. Subsequently, in Sec. 5, we describe the proposed DDSB. In addition, we discuss in Sec. 6 the conditions for the DDSB to convergence to the optimal centralized beamformer and investigate its convergence rate. In Sec. 7, we provide some simulation results. Finally, in Sec. 8, conclusions are drawn.

2. NOTATION

We consider a randomly connected WSN with N nodes, where each node i computes the noisy speech discrete Fourier transform (DFT) coefficients $Y_i(k,m)$ on a frame-by-frame basis, where k and m denote the frequency bin and time frame index, respectively. The DFT coefficients are assumed to be random variables, indicated by upper case letters, while their realizations will be indicated by lower case letters. We assume an additive noise model, i.e.,

$$Y_i(k,m) = X_i(k,m) + V_i(k,m), \tag{1}$$

where $X_i(k,m)$ and $V_i(k,m)$ indicate the speech and noise DFT coefficients at microphone (node) i, respectively. Further, we assume

This research is partly supported by the Dutch Technology Foundation STW.

that the speech and noise are uncorrelated, i.e.,

$$E\left[X_i(k,m)V_j(k,m)\right] = 0 \quad \forall \quad k,m,i,j, \tag{2}$$

where $E[\cdot]$ denotes the statistical expectation operator. The speech and noise DFT coefficients are assumed to be independent across time and frequency, which allows us to omit the time and frequency indices for notational convenience. We use a stacked vector notation $\mathbf{Y} = [Y_1, \cdots, Y_N]^T$ consisting of Y_i for all nodes i, where the superscript T denotes transposition of a vector or a matrix. The speech and noise vector $\mathbf{X}$ and $\mathbf{V}$ are defined similarly as $\mathbf{Y}$. The WSN data model for all nodes can then be written as $\mathbf{Y} = \mathbf{X} + \mathbf{V}$. In this paper, we consider the case of a single desired speech source. The speech vector can then be written as

$$\mathbf{X} = \mathbf{d}S, \tag{3}$$

where S is a desired source DFT coefficient, and $\mathbf{d}$ models the acoustic transfer function from the speech source S to all sensor nodes.

3. CENTRALIZED BEAMFORMER

A centralized beamformer aims to estimate the desired speech DFT coefficient S by computing a weighted linear combination of the elements in $\mathbf{Y}$, i.e.,

$$Z = \mathbf{w}^H \mathbf{Y}, \tag{4}$$

where Z is the estimated clean speech DFT coefficient, $\mathbf{w}$ is a vector with filter coefficients and $(\cdot)^H$ denotes Hermitian transposition.

The optimal filter $\mathbf{w}$ that minimizes the contribution of interferences to the output Z subject to the constraint of no speech distortion can be computed as

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R_{YY}} \mathbf{w}, \text{ subject to } \mathbf{w}^H \mathbf{d} = 1, \tag{5}$$

where $\mathbf{R_{YY}}$ is defined as $\mathbf{R_{YY}} = E[\mathbf{Y}\mathbf{Y}^H]$. Assuming that V_i, $\forall i$ are zero mean, spatially uncorrelated with power spectral density (PSD) $\sigma^2_{V_i}$ and uncorrelated with the speech source, $\mathbf{R_{VV}} = \text{diag}\left\{\sigma^2_{V_1}, \cdots, \sigma^2_{V_N}\right\}$ and $\mathbf{R_{YY}} = \mathbf{R_{XX}} + \mathbf{R_{VV}}$. Using the matrix inversion lemma [5], the filter $\mathbf{w}$ that solves the above constrained optimization problem (5) is given by

$$\mathbf{w} = \frac{\mathbf{R}^{-1}_{\mathbf{VV}}\mathbf{d}}{\mathbf{d}^H \mathbf{R}^{-1}_{\mathbf{VV}}\mathbf{d}}. \tag{6}$$

The above made assumption of spatially uncorrelated noise is validated for a diffuse noise field and/or when the distance between sensors is sufficiently large. Notice that this version of the delay and sum beamformer is more general than the generally used delay and sum beamformer, as also presented in [5], as it allows for different noise PSDs per microphone. The acoustic transfer functions $\mathbf{d}$ can be uniquely determined by gain and delay values as $\mathbf{d} = \left[a_1 e^{-j\omega_k \tau_1}, \cdots, a_N e^{-j\omega_k \tau_N}\right]^T$, where a_i is the damping coefficient, and τ_i denotes the delay in number of samples. The beamformer output is then given by

$$Z = \frac{\sum_{i=1}^N a_i (\sigma^2_{V_i})^{-1} e^{j\omega_k \tau_i} Y_i}{\sum_{i=1}^N a_i^2 (\sigma^2_{V_i})^{-1}}. \tag{7}$$

To obtain the optimal central solution (7), the centralized beamformer requires that there is a central processing unit that has knowledge of all sensor positions (at least relative to the sources) and each sensor has to broadcast its DFT coefficient Y_i to the central processor. However, the centralized processing will neither be robust nor scalable when the size of the WSN grows.

4. RANDOMIZED GOSSIP ALGORITHM

In order to guide the reader and help to appreciate the contribution of this paper, we give in this section a brief overview of the most essential aspects of the randomized gossip algorithm [4].

The randomized gossip algorithm can be used to solve consensus problems in distributed way. Given a randomly connected network of N nodes and initial scalar value $g_i(0)$ at node i, the randomized gossip algorithm aims to find the average value $g_{ave} = \frac{1}{N}\sum_{i=1}^N g_i(0)$ at all nodes by using an iterative scheme and using only local information and local processing. Let $\mathbf{g}(t) = [g_1(t), \cdots, g_N(t)]^T$ denote the vector of values at the end of iteration t. At iteration t, each node in the asynchronous gossip runs an independent Poisson clock. When node i's clock ticks, it randomly selects and communicates with one neighboring node j with probability p_{ij}. All probabilities p_{ij} can be stacked in an $N \times N$ probability matrix $\mathbf{p}$, with $p_{ij} > 0$ if there is a communication link between node i and node j, otherwise $p_{ij} = 0$. In each iteration, a node i and a node j exchange their local information and update their current local estimates as $g_i(t) = g_j(t) = (g_i(t-1) + g_j(t-1))/2$. Except these two active nodes, other inactive nodes in the network keep the same estimates as during the last iteration $t-1$. A general vector formulation of the gossip algorithm is given by

$$\mathbf{g}(t) = \mathbf{U}(t)\mathbf{g}(t-1), \tag{8}$$

where $\mathbf{U}(t)$ is a randomly selected $N \times N$ dimensional update matrix, which is selected independently across time. For two communicating nodes i and j at iteration t, the update matrix $\mathbf{U}(t)$ is

$$\mathbf{U}(t) = \mathbf{I} - \frac{1}{2}\left(e_i - e_j\right)\left(e_i - e_j\right)^T, \tag{9}$$

where $e_i = [0, \cdots, 0, 1, 0, \cdots, 0]^T$ is an N dimensional vector with the ith component equal to 1. When $\mathbf{U}(t)$ is a doubly stochastic matrix and the network is connected, all nodes in the network are guaranteed to converge to the average value g_{ave}. We will give a short summary of the convergence conditions and convergence rate in Sec. 6, while the detailed proof is given in [4].

5. DISTRIBUTED DELAY AND SUM BEAMFORMER

Unlike the centralized beamformer, the proposed distributed beamformer broadcasts information of a node i only to one of its neighbors, and aims to obtain the same optimal estimated signal as (7) by using only local information and local processing.

We assume that each node i in the WSN for a given time frame has two initial values $\widetilde{Y}_i(0) = d_i^H(\sigma^2_{V_i})^{-1}Y_i = a_i(\sigma^2_{V_i})^{-1}e^{j\omega\tau_i}Y_i$ and $\widetilde{d}_i(0) = d_i^H(\sigma^2_{V_i})^{-1}d_i = a_i^2(\sigma^2_{V_i})^{-1}$, where realizations of Y_i are obtained using the microphone at the node i, and d_i and $\sigma^2_{V_i}$ have to be estimated. In this paper we assume d_i is known and estimate $\sigma^2_{V_i}$ using an ideal voice activity detector in order to focus on the distributed beamforming algorithm. In practice, d_i can be estimated and adapted using [6] and the noise PSD $\sigma^2_{V_i}$ can be estimated using e.g., [7]. Let $\widetilde{\mathbf{Y}}(0)$ be a stacked N dimensional vector defined as $\widetilde{\mathbf{Y}}(0) = [\widetilde{Y_1}(0), \cdots, \widetilde{Y_N}(0)]^T$, similarly, all $\widetilde{d}_i(0)$ are stacked in an N dimensional vector $\widetilde{\mathbf{d}}(0)$.

The optimal centralized estimates (7) can be obtained as

$$Z = \widetilde{Y}_{ave}/\widetilde{d}_{ave}, \tag{10}$$

where $\widetilde{Y}_{ave} = \frac{1}{N}\mathbf{1}^T\widetilde{\mathbf{Y}}(0)$ and $\widetilde{d}_{ave} = \frac{1}{N}\mathbf{1}^T\widetilde{\mathbf{d}}(0)$ with $\mathbf{1}$ denoting the vector of all ones. Then the goal of the algorithm introduced in this paper is to find the average value $\widetilde{Y}_{ave}$ and $\widetilde{d}_{ave}$ in a distributed way. Based on gossip processing, this algorithm is an iterative scheme referred to as DDSB. The iterative DDSB scheme considered here is randomized and asynchronous, since at each iteration a random pair of nodes is active.

Let $\widetilde{\mathbf{Y}}(t)$ and $\widetilde{\mathbf{d}}(t)$ be defined as vector $\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{d}}$ at iteration t, respectively. A general vector form of the DDSB which describes the current estimate for each iteration t is given by

$$\widetilde{\mathbf{Y}}(t) = \mathbf{U}(t)\widetilde{\mathbf{Y}}(t-1), \tag{11}$$

$$\widetilde{\mathbf{d}}(t) = \mathbf{U}(t)\widetilde{\mathbf{d}}(t-1), \tag{12}$$

$$\widetilde{Z}_i(t) = \widetilde{Y}_i(t)/\widetilde{d}_i(t), \tag{13}$$

with $\widetilde{Z}_i(t)$ the DDSB output at iteration t.

6. CONVERGENCE

The convergence of $\lim_{t\to\infty}\widetilde{\mathbf{Y}}(t)$ to $\widetilde{Y}_{ave}\mathbf{1}$ and $\lim_{t\to\infty}\widetilde{\mathbf{d}}(t)$ to $\widetilde{d}_{ave}\mathbf{1}$ is guaranteed for any initial vector $\widetilde{\mathbf{Y}}(0)$ and $\widetilde{\mathbf{d}}(0)$ as long as the update matrix is a doubly stochastic matrix [4]. Since $\widetilde{\mathbf{Y}}(t)$ and $\widetilde{\mathbf{d}}(t)$ converge, also their ratio converge as long as $\widetilde{d}_{ave} \neq 0$, i.e., $\widetilde{Z}_i(t) = \frac{\lim_{t\to\infty}\widetilde{Y}_i(t)}{\lim_{t\to\infty}\widetilde{d}_i(t)} = Z$.

Let the convergence error be defined as $CE = \frac{\left\|\widetilde{\mathbf{Y}}(t)-\widetilde{Y}_{ave}\mathbf{1}\right\|}{\left\|\widetilde{\mathbf{Y}}(0)\right\|}$. The convergence rate of the algorithm can then in analogy with [4] be defined by the ϵ-averaging time $T_{ave}(\epsilon, P)$ as

$$\sup_{\widetilde{\mathbf{Y}}(0)} \inf_{t=0,1,\cdots} \left\{P\left(CE \geq \epsilon\right) \leq \epsilon\right\}, \tag{14}$$

and can be shown to be bounded by the second largest eigenvalue of the expected value of the update matrix $E\left[\mathbf{U}\right]$. That is [4],

$$\frac{0.5\log\epsilon^{-1}}{-\log\lambda_2\left(E\left[\mathbf{U}\right]\right)} \leq T_{ave}(\epsilon, E\left[\mathbf{U}\right]) \leq \frac{3\log\epsilon^{-1}}{-\log\lambda_2\left(E\left[\mathbf{U}\right]\right)}. \tag{15}$$

The smaller the magnitude of the eigenvalue $\lambda_2\left(E\left[\mathbf{U}\right]\right)$, the faster the convergence will become. For a doubly stochastic probability matrix $\mathbf{p}$, it can be shown that [4]

$$E\left[\mathbf{U}\right] = \left(1 - 1/N\right)\mathbf{I} + \mathbf{r}/N, \tag{16}$$

with $\mathbf{r} = (\mathbf{p}+\mathbf{p}^T)/2$ and $\mathbf{p}$ as definded in Sec. 4. From Eq. (15) and (16) we see that $\lambda_2\left(E\left[\mathbf{U}\right]\right)$, and thus the averaging time, depends on $\mathbf{p}$ and thus on the underlying network topology.

In practice, the exact topology is generally unknown. For the case of a fully connected network and uniform pairwise gossiping, it was mentioned in [8], that T_{ave} has an upper bound, say $T_{ave,FC}$, which equals $T_{ave} \leq T_{ave,FC} = \frac{3\log\epsilon^{-1}}{-\log(1-1/N)}$. This is thus an upper bound for T_{ave} for the best connected network. To be more specific about the averaging time of the proposed algorithm expressed in terms of sensors in the network, we will in this section derive an upper bound for the averaging time for the worst connected network.

To derive such a bound, we constrain matrix $\mathbf{p}$ to be doubly stochastic. In that case, the worst possible connected network is given by a set of sensors that are connected in a closed circle, where the probability that a sensor connects to the next (clockwise) sensor is denoted by q, while the probability that it connects to the previous (anti-clockwise) sensor is denoted by $1-q$. This leads to the following $\mathbf{p}$ matrix,

$$\mathbf{p} = \begin{bmatrix} 0 & q & & & & 1-q \\ 1-q & 0 & q & & 0 & \\ & 1-q & \ddots & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & 0 & & \ddots & \ddots & q \\ q & & & & 1-q & 0 \end{bmatrix}. \tag{17}$$

For this $\mathbf{p}$ matrix, the $\mathbf{r}$ matrix in Eq. (16) is also doubly stochastic with a similar structure as Eq. (17), but then with $q = 0.5$.

To obtain a bound of T_{ave} for this network, say $T_{ave,WC}$, we need to find the second largest eigenvalue of the matrix expressed by Eq. (16). Since $\mathbf{I}(1-\frac{1}{N})$ is an identity matrix, it is sufficient to find the second largest eigenvalue of $\mathbf{r} = (\mathbf{p}+\mathbf{p}^T)/2$ with $\mathbf{p}$ as in (17). Matrix $\mathbf{r}$ is known as a scaled Gear-matrix [9]. The eigenvalues of such a matrix have a special form given by [9] $\lambda_i = 2\beta\cos(2\pi n/N)$, with $n \in \{0,...,N-1\}$, with scaling $\beta = 0.5$.

The second largest eigenvalue of E$\{\mathbf{U}\}$ is thus explicitly given by $\lambda_2 = 1 - \frac{1}{N} + \frac{1}{N}\cos(2\pi/N)$. Using Eq. (15) this leads to the upper bound

$$T_{ave} \leq T_{ave,WC} = \frac{3\log\epsilon^{-1}}{-\log\left\{1-\frac{1}{N}(1-\cos(2\pi/N))\right\}} \tag{18}$$

$$\leq \frac{3N^3\log\epsilon^{-1}}{(2\pi)^2/2 + \sum_{k=2}^{\infty}(-1)^{k+1}\frac{N^2}{(2k)!}\left(\frac{2\pi}{N}\right)^{2k}}, \tag{19}$$

where, for going from (18) to (19) we made use of the truncated Taylor series expansion $\log\{1-x\} = -\sum_{k=1}^{\infty}\frac{x^k}{k}$ for $-1 \leq x < 1$ and $\cos x = \sum_{k=0}^{\infty}(-1)^k\frac{x^{2k}}{(2k)!}$ [10], to write $T_{ave,WC}$ explicitly in terms of the number of sensors N.

In conclusion, in case of the worst connected graph, the averaging-time grows thus in the order $\mathcal{O}(N^3)$ with the number of sensors N. However, in many practical situations, the network graph will be better connected than the worst case scenario considered here as we will show in Sec. 7. Notice, that for a fully connected network, T_{ave} grows in the order $\mathcal{O}(N)$ [8].

7. SIMULATIONS

We simulated a WSN where $N = 20$ microphones, a speech source and a noise source are randomly distributed in a 10×10 room. The 20 microphones are randomly connected with 60 edges, and gather noisy speech signals at a sampling frequency of $f_s = 16$kHz. The desired source signal is a 30 sec. speech signal [11], and the interfering noise source is a zero-mean white Gaussian point source. All nodes process the signals frame by frame, with a frame length of 25ms and Hann window. We assume that the distance l_i between node i and the desired source is known leading to damping $a_i = 1/l_i$ and delay $\tau_i = \frac{l_i}{c}f_s$ with the speed of sound $c = 340\frac{m}{s}$. For the DDSB, we use a fixed number of iteration per time frame. The noisy signal per microphone is simulated according to (1) by properly delaying and scaling the target and interfering signal. The optimal probability vector $\mathbf{p}_i$, i.e., the ith row of $\mathbf{p}$ is computed using only local communication as described in [4]. As a measure to access the performance of the DDSB, the mean square error (MSE) for node i and time-frame m is defined as

$$\text{MSE}_i(m) = \frac{1}{K}\sum_{k=1}^{K}\left\|\widehat{Z}_i(k,m) - S(k,m)\right\|^2, \tag{20}$$

where K is the number of frequency bins, and $\widehat{Z_i}(k,m)$ is a DFT coefficient of the beamformer output, i.e., DDSB or CDSB. The MSE averaged over all time frames is then given by $\text{MSE}_i^* = \frac{1}{M}\sum_{m=1}^{M} \text{MSE}_i(m)$, with M the number of time frames.

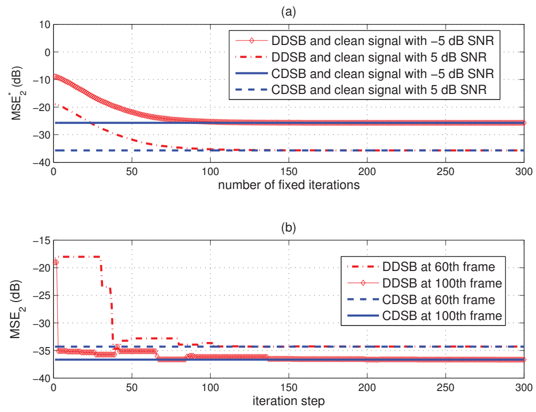

Fig. 1. (a) The MSE of node 2 versus the number of iterations. (b) The MSE of node 2 with 5 dB input SNR versus iteration step.

Fig. 1(a) shows the MSE between the DDSB output of node 2 and the clean speech signal, compared to the MSE between the CDSB output and the clean speech signal. As expected, the MSE in different SNR situations decreases with increasing fixed number of iterations. We observed that, when each pair of neighboring nodes communicates frequently enough, the DDSB reaches the same performance as the CDSB. In Fig. 1(b), we show two examples of per frame convergence in terms of the MSE between the beamformer output and the desired speech signal versus the iteration number. Without surprise, the MSE decreases with increasing iterations for all frames. We see indeed that the DDSB converges asymptotically to the CDSB.

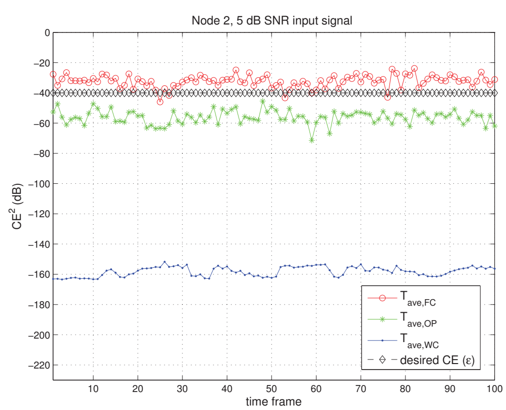

Fig. 2. Convergence error CE across time frames.

Next, we compare the convergence error CE for different fixed number of iterations, to compare how far CE is from ϵ. At first, we use $T_{ave,OP}$ which is based on the upper bound in (15) in combination with the optimal $\mathbf{p}$ matrix [4]. We also compare this to the upper bound of the fully connected network $T_{ave,FC}$ and the upper bound of the worst connected network $T_{ave,WC}$. The result is shown in Fig. 2. In addition we show the desired CE for $\epsilon = 0.01$. It shows that with both $T_{ave,WC}$ and the optimal $T_{ave,OP}$, lower convergence error than the desired CE is obtained, and that with $T_{ave,FC}$ higher CE than the desired CE is obtained. Given ϵ, we see that $T_{ave,OP}$ is the least number of iterations to guarantee convergence ϵ for a given connected network, and $T_{ave,WC}$ is the least number of iterations to guarantee convergence ϵ given only the network size N.

8. CONCLUSIONS

In this paper, we proposed a distributed delay and sum beamformer (DDSB), an algorithm for decentralized estimation of the clean speech signal in a wireless sensor network. The algorithm has no topology constraints and uses only local information exchange and local processing, while its output converges asymptotically to the optimal centralized beamformer. The convergence rate of the DDSB is inversely proportional to the second largest eigenvalue of the expected value of the update matrix. Moreover, we described an upper convergence bound of the DDSB for a given size network. Finally, we provided simulation results to show the effectiveness of our algorithm and to show that it converges to the centralized beamformer.

9. REFERENCES

[1] A. Bertrand and M. Moonen, "Distributed adaptive node-specific signal estimation in fully connected sensor networks – part II: Simultaneous and asynchronous node updating," *IEEE Trans. Signal Processing*, vol. 58, no. 10, pp. 5292 –5306, oct. 2010.

[2] A. Bertrand and M. Moonen, "Distributed adaptive estimation of node-specific signals in wireless sensor networks with a tree topology," *IEEE Trans. Signal Processing*, vol. 59, no. 5, pp. 2196–2210, May 2011.

[3] Y. Hioka and W.B. Kleijn, "Distributed blind source separation with an application to audio signals," in *IEEE Int. Conf. Acoust., Speech, Signal Processing*, may 2011, pp. 233 –236.

[4] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Trans. on Information Theory*, vol. 52, no. 6, pp. 2508 – 2530, june 2006.

[5] M. Brandstein and D. Ward (Eds.), *Microphone arrays*, Springer, 2001.

[6] 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, 2001.

[7] R. C. Hendriks, R. Heusdens, and J. Jensen, "MMSE based noise PSD tracking with low complexity," in *IEEE Int. Conf. Acoust, Speech, Signal Processing*, 2010, pp. 4266–4269.

[8] A.G. Dimakis, S. Kar, J.M.F. Moura, M.G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847 –1864, nov. 2010.

[9] C. W. Gear, "A simple set of test matrices for eigenvalue programs," *Mathematics of Computation*, vol. 23, no. 105, pp. pp. 119–125, 1969.

[10] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series and Products*, New York: Academic, 6th ed. edition, 2000.

[11] J. S. Garofolo, "DARPA TIMIT acoustic-phonetic speech database," *National Institute of Standards and Technology (NIST)*, 1988.