UTILITY BASED CROSS-LAYER COLLABORATION FOR SPEECH ENHANCEMENT IN WIRELESS ACOUSTIC SENSOR NETWORKS

Joseph Szurley, Alexander Bertrand*, Marc Moonen*, Peter Ruckebusch†, and Ingrid Moerman†*

* Electrical Engineering Dept. (ESAT-SCD)
Katholieke Universiteit Leuven
Kasteelpark Arenberg 10
B-3001 Leuven, Belgium
E-mail: joseph.szurley@esat.kuleuven.be,
alexander.bertrand@esat.kuleuven.be,
marc.moonen@esat.kuleuven.be

† IBBT, Department of Information Technology (INTEC)
Ghent University
Gaston Crommenlaan 8, Bus 201
9050 Ghent, Belgium
E-mail: peter.ruckebusch@intec.ugent.be,
ingrid.moerman@intec.ugent.be

ABSTRACT

A wireless acoustic sensor network is considered that is used to estimate a desired speech signal that has been corrupted by noise. The application layer of the WASN derives an optimal filter in a linear MMSE sense. A utility function is then used in conjunction with the MMSE estimate in order to evaluate the most significant signal components from each node in the system. The utility values are used as a cross-layer link between the application layer and the network layer so the nodes transmit the signal components that are deemed most relevant to the estimate while adhering to the power constraints of the system. The simulation results show that a high signal-to-error and signal-to-noise ratio is still achievable while transmitting a subset of signal components.

1. INTRODUCTION

Wireless sensor networks (WSN)s have seen phenomenal interest in recent years pertaining to algorithm development and deployment. This interest can be attributed to many attractive and unique factors of WSNs such as relatively low power consumption on a per node basis, the ability to monitor large areas at a relatively low cost of deployment, and the resilience to failure of individual nodes [1, 2]. The nodes of the WSN work cooperatively to collect, estimate, and transmit data and may do this in a distributed or centralized fashion. In a centralized WSN a fusion center collects data from the nodes and can perform the majority of the processing. The nodes in this scenario are simply used for data collection, transmission, and in some cases also for data compression.

Recent research has started to exploit the versatility of WSNs with applications relating to acoustic signals [3–5]. Wireless acoustic sensor networks (WASN)s differ from most other WSNs in that the signals observed often have characteristics that fluctuate much quicker in time which require significantly faster transmission and computation speeds. A direct consequence of an increase in transmission rates is a higher power consumption of the nodes. Depending on the application, for the three main functions of a node in a distributed WASN, collection, estimation, and transmission nearly 80% of the power consumed is used in the latter [6]. In a centralized WASN the bulk of node power is used to transmit data to the fusion center.

This research work was carried out at the ESAT Laboratory of Katholieke Universiteit Leuven, in the frame of K.U.Leuven Research Council CoE EF/05/006 'Optimization in Engineering' (OPTEC) and PFV/10/002 (OPTEC), Concerted Research Action GOA-MaNet, the Belgian Programme on Interuniversity Attraction Poles initiated by the Belgian Federal Science Policy Office IUAP P6/04 'Dynamical systems, control and optimization' (DYSCO) 2007-2011, Research Project FWO nr. G.0600.08 'Signal processing and network design for wireless acoustic sensor networks'. Alexander Bertrand is supported by a Ph.D grant of the I.W.T. (Flemish Institute for the Promotion of Innovation through Science and Technology). The scientific responsibility is assumed by its authors.

One strategy to alleviate the transmission burden on the individual nodes is to reduce the amount of data used in the estimation process at the fusion center. This places added burden on the node in order to process and compress the signal [6, 7]. The effect on the estimation process must also be taken into account due to the fact that only a truncated or compressed version of the signal is sent to the fusion center. The effects of compression or rate-constrained transmissions of audio signals have been studied in [6, 8].

The fusion center in a centralized WASN uses an application layer for signal estimation and a network layer to manage communications between the individual nodes. In WASNs there is often little interaction between the two layers as both function independently on their primary tasks e.g. audio compression to alleviate transmission burden, happens independently from the estimation process. It is therefore useful to initiate cross-layer communication so that relevant information from the application layer can be used to help facilitate better network management.

In this paper a centralized WASN is explored where the fusion center not only uses a MMSE technique to reconstruct a desired speech signal that has been corrupted by noise but also uses the information from the estimation as a network management tool. This information is then passed to the individual nodes so that data transmission utilizes the transmit power in the most efficient fashion. The main motivation behind this paper is to decrease the amount of power needed for transmission while still accurately being able to preserve the intelligibility of the speech signal estimate. The reduced transmit power then leads to reduced power consumption on an individual node basis and therefore over the system as a whole.

The paper is organized as follows. Section 2 covers linear MMSE signal estimation with optimal adaptive filtering. Section 3 reviews the calculation of the utility from the MMSE estimates. Section 4 discusses the cross-layer communications between the application layer and the network layer. Section 5 shows simulation results for the WASN when the utility is used. Section 6 draws conclusions from the simulation and discusses their implications on cross-layer node-fusion center collaboration performance.

2. MINIMUM MEAN SQUARE ERROR ESTIMATION

In the envisaged WASN there are N nodes each with one microphone. This formulation can be extended to nodes with multiple microphones in a similar fashion as in [3, 4]. Node k collects a speech signal corrupted by an uncorrelated zero mean noise. The microphone signals at node k are then given in the short-time Fourier transform (STFT) domain by

$$y_k(\omega_l, \mathrm{t}) = d_k(\omega_l, \mathrm{t}) + v_k(\omega_l, \mathrm{t}) \quad (1)$$

where d_k is the desired speech, v_k is the added noise, ω_l is the discrete frequency with $l \in \{1, ..., L\}$, and t is the frame index where t $\in \mathbb{N}$. The received signals are stacked into a vector $\mathbf{y}$ which takes

© EURASIP, 2011 - ISSN 2076-1465

the form of

$$\mathbf{y}[\omega_l,\mathrm{t}] = [y_1[\omega_l,\mathrm{t}],...,y_k[\omega_l,\mathrm{t}],...,y_N[\omega_l,\mathrm{t}]]^T. \quad (2)$$

In the sequel, we will often omit the frequency index ω_l and the frame index t bearing in mind that the operations are taking place in the STFT domain. It is assumed that the signals are short term ergodic and change slowly over time allowing for a STFT representation.

The goal of the linear minimum mean square error (MMSE) estimation is to minimize the difference between an observed signal d and a linearly filtered version of the sensor signal $\hat{d} = \hat{\mathbf{w}}^H\mathbf{y}$, where the superscript ^{H} denotes the conjugate transpose. The signal d is designated as the speech component of a reference microphone. It is assumed that the reference microphone, y_1, is a microphone at the fusion center.

The linear MMSE estimation may be realized in terms of a mean square error (MSE) cost function represented by

$$J_{MSE}(\mathbf{w}) = \mathscr{E}\{\|d - \mathbf{w}^H\mathbf{y}\|^2\} \quad (3)$$

where $\mathscr{E}$ is the expectation operator. Minimizing the cost function in terms of $\mathbf{w}$ gives the linear MMSE estimator which takes the form of the well known multi-channel Wiener filter (MWF) [9, 10]

$$\hat{\mathbf{w}} = \mathbf{R}_{\mathbf{yy}}^{-1}\mathbf{r}_{\mathbf{y}d} \quad (4)$$

where the auto-correlation matrix is defined as $\mathbf{R}_{yy} = \mathscr{E}\{\mathbf{y}\mathbf{y}^H\}$, the cross-correlation vector is defined as $\mathbf{r}_{\mathbf{y}d} = \mathscr{E}\{\mathbf{y}d^*\}$, and d^* indicates the complex conjugate of the signal d. The MWF may be evaluated on an individual frequency bin basis which corresponds to multi-tap filtering in the time domain.

The speech reference signal, d, used in (3) is normally unknown but $\mathbf{r}_{\mathbf{y}d}$ can be estimated using the auto-correlation matrices of the noise and speech-plus-noise signals [4, 5, 9]

$$\mathbf{R}_{\mathbf{dd}} = \mathbf{R}_{\mathbf{yy}} - \mathbf{R}_{\mathbf{vv}} \quad (5)$$

where $\mathbf{R}_{\mathbf{vv}} = \mathscr{E}\{\mathbf{v}\mathbf{v}^H\}$, $\mathbf{R}_{\mathbf{dd}} = \mathscr{E}\{\mathbf{d}\mathbf{d}^H\}$, and $\mathbf{y} = \mathbf{d} + \mathbf{v}$ by combining (1) and (2). If a perfect voice activity detection (VAD) is assumed then $\mathbf{R}_{\mathbf{yy}}$ can be estimated during speech-plus-noise periods and $\mathbf{R}_{\mathbf{vv}}$ can be estimated during noise only periods. By multiplying $\mathbf{R}_{\mathbf{dd}}$ by a unit vector $\mathbf{e}_d$ that is equal to 1 corresponding to the index of the reference microphone and zero elsewhere, the cross-correlation vector $\mathbf{r}_{\mathbf{y}d}$ can be extracted.

The signal correlation matrices are often updated at discrete time intervals t by means of a forgetting factor $0 < \lambda < 1$

$$\mathbf{R}_{\mathbf{yy}}[\omega_l,\mathrm{t}] = \lambda\mathbf{R}_{\mathbf{yy}}[\omega_l,\mathrm{t}-1] + (1-\lambda)\mathbf{y}[\omega_l,\mathrm{t}]\mathbf{y}[\omega_l,\mathrm{t}]^H. \quad (6)$$

This allows the correlation matrix to use current data as well as long time averaged statistics.

3. UTILITY FUNCTION

In [11] a utility function was introduced to monitor the MSE cost increase when deleting a *node* from the system or the MSE cost decrease when adding a *node* to the system. The utility function offers an efficient way to determine not only how much each node is contributing to the system as a whole but more importantly how the individual signal characteristics per node, i.e. per frequency bin, effect the signal estimation. This is accomplished by monitoring the change of the cost function of the system with respect to the addition and deletion of individual *microphones per frequency*. The information from the utility function can then be passed on to the network layer of each individual node in order to best utilize the available transmit power.

Before outlining the equations for microphone deletion and microphone addition it will be constructive to place the microphones per frequency into two groups, an active group $\mathbb{A}(\omega_l)$ which contribute to the current signal estimation and an inactive group $\mathbb{I}(\omega_l)$ which are not used during signal transmission from the nodes to the fusion center. The microphones in the system can thus represented as the union of the active and inactive sets $\mathbb{A} \cup \mathbb{I}$.

This notation is also extended to the received signal as $\mathbf{y}_{\mathbb{A}}$ and $\mathbf{y}_{\mathbb{I}}$ indicating signals that are used for the current estimation and signals that are not respectively. Similarly the correlation matrices, correlation vectors, and corresponding filters all have active and inactive components.

3.1 Microphone Deletion

For microphone deletion, the utility for a given microphone k is evaluated by comparing the MSE cost function before and after deleting the microphone. The utility therefore determines how strongly the current estimate is effected by the exclusion of microphone k.

The cost function (3) with a microphone k removed, $\forall k \in \mathbb{A}(\omega)$, is defined as

$$J_{\mathbb{A}-k}(\mathbf{w}_{\mathbb{A}-k}) = \mathscr{E}\{\|d - \mathbf{w}_{\mathbb{A}-k}^H\mathbf{y}_{\mathbb{A}-k}\|^2\} \quad (7)$$

where $\mathbf{y}_{\mathbb{A}-k}$ is the received signal without the microphone k. The utility U_k is defined then as the difference between the optimal values of the cost function with and without the microphone signal per frequency and is given by

$$U_k = J_{\mathbb{A}-k}(\hat{\mathbf{w}}_{\mathbb{A}-k}) - J_{\mathbb{A}}(\hat{\mathbf{w}}_{\mathbb{A}}) \quad (8)$$

where $\hat{\mathbf{w}}_{\mathbb{A}}$ and $\hat{\mathbf{w}}_{\mathbb{A}-k}$ denote the optimal filter and the optimal filter without microphone k, respectively.

Only the active microphone set is used to calculate the utility as the inactive set does not contribute to the current signal estimation. Notice that microphones that contribute very little to the cost function will have a very low utility, conversely microphones that have a very high contribution will have larger utility values.

The utility for each active microphone U_k is placed in a vector taking the form

$$\mathbf{u}_{\mathbb{A}} = [U_1...U_k]^T \quad (9)$$

which monitors all active microphones simultaneously in the given frequency bin. The microphones can then be ranked by their importance to the current estimation of the signal. The utility after some algebratic manipulation was shown to be able to be efficiently calculated in [11] by

$$\mathbf{u}_{\mathbb{A}} = \mathbf{\Lambda}^{-\mathbf{1}} \mid \hat{\mathbf{w}}_{\mathbb{A}} \mid^2 \quad (10)$$

where $\mathbf{\Lambda}_{i,j}$ monitors the diagonal elements of the inverse correlation matrix,

$$[\mathbf{\Lambda}]_{i,j} = [\mathbf{R}_{\mathbf{y}_{\mathbb{A}}\mathbf{y}_{\mathbb{A}}}^{-1}]_{i,j}\delta_{i,j} \quad (11)$$

and $\delta_{i,j}$ is the Kronecker delta function

$$\delta_{i,j} = \begin{cases} 1 & \text{for i = j} \\ 0 & \text{for i} \neq \text{j} \end{cases}. \quad (12)$$

Since $\mathbf{R}_{\mathbf{y}_{\mathbb{A}}\mathbf{y}_{\mathbb{A}}}^{-1}$ is already available from the computation of $\hat{\mathbf{w}}_{\mathbb{A}}$ (see (4)), there are no extra computations involved. Since $\mathbf{\Lambda}$ is a diagonal matrix, equation (10) can be computed at almost no extra computational cost.

3.2 Microphone Addition

Much like microphone deletion, microphone addition aims to determine which microphones in the inactive set would contribute the most to the current estimation. The cost function with a microphone k added, $k \in \mathbb{I}(\omega)$, takes the form

$$J_{\mathbb{A}+k}(\mathbf{w}_{\mathbb{A}+k}) = \mathscr{E}\{\|d - \mathbf{w}_{\mathbb{A}+k}^H\mathbf{y}_{\mathbb{A}+k}\|^2\} \quad (13)$$

where $\mathbf{y}_{\mathbb{A}+k}$ is the received signal with microphone k added.

The utility is then given by the difference between the optimal values of the current cost function and the cost function with an added microphone

$$U_k = J_{\mathbb{A}}(\hat{\mathbf{w}}_{\mathbb{A}}) - J_{\mathbb{A}+k}(\hat{\mathbf{w}}_{\mathbb{A}+k}) \tag{14}$$

where $\hat{\mathbf{w}}_{\mathbb{A}}$ is defined in (8) and $\hat{\mathbf{w}}_{\mathbb{A}+k}$ is denoted as the optimal filter with microphone k added.

In order to evaluate the utility for microphone addition it will be constructive to define two intermediate variables as in [11]

$$[\mathbf{\Sigma}]_{i,j} = [\mathbf{R}_{\mathbf{y}_{\mathbb{I}}\mathbf{y}_{\mathbb{I}}}]_{i,j}\delta_{i,j} - [\mathbf{R}^{H}_{\mathbf{y}_{\mathbb{A}}\mathbf{y}_{\mathbb{I}}}]_{i,j}\delta_{i,j} \tag{15}$$

which is a diagonal matrix and

$$\mathbf{V} = \mathbf{R}^{-1}_{\mathbf{y}_{\mathbb{A}}\mathbf{y}_{\mathbb{A}}}\mathbf{R}_{\mathbf{y}_{\mathbb{A}}\mathbf{y}_{\mathbb{I}}}. \tag{16}$$

Note that in (16), $\mathbf{R}^{-1}_{\mathbf{y}_{\mathbb{A}}\mathbf{y}_{\mathbb{A}}}$ is already evaluated from the estimation procedure in (4) so no extra matrix inversion is required.

These intermediate variables monitor the correlation statistics between the microphones that are in the active set, hence used for the current signal estimation, and microphones that are in the inactive set, which are not used in the transmission signal from the nodes.

The two intermediate variables (15,16) are then used to estimate the utility for microphone addition in the form

$$\mathbf{u}_{\mathbb{I}} = \mathbf{\Sigma}^{-1}|\mathbf{V}^{\mathrm{T}} r^{*}_{\mathbf{y}_{\mathbb{A}}d} - r_{\mathbf{y}_{\mathbb{I}}d}|^2 \tag{17}$$

where $\mathbf{r}_{\mathbf{y}_{\mathbb{A}}d}$ and $\mathbf{r}_{\mathbf{y}_{\mathbb{I}}d}$ are the cross-correlations between the active and inactive microphones and the desired signal respectively, and the superscript ^{T} is the transpose operator.

4. CROSS-LAYER COLLABORATION

The utility function is a feasible and efficient method to rank the microphones in a WASN to the overall benefit of the signal estimation. These values can be used in conjunction with the network layer in dictating power allocation for the transmission of signals from the nodes. The fusion center first collects data from the individual nodes and estimates the signal using the MSE criterion. It then decides which microphones from the individual nodes influence the estimation procedure in a desired frequency bin the most. This information is then relayed to the network layer and the corresponding nodes. On the next update period, for each node, only a subset of the frequency bins are transmitted and used in the estimation.

The question then becomes which is the best subset of the full signal for the estimation procedure. The utility values may be exploited in several different ways to benefit the estimation while adhering to the transmit power constraints. In subsections 4.1 and 4.2 we describe the greedy algorithm which is shown as the pseudocode **Algorithm 1**. This algorithm is used for the simulations in section 5.

4.1 Utility values and cross-layer collaboration

The way the utility is used to determine the appropriate signal characteristics has a profound impact on how the signal estimation algorithm proceeds. The system now is not only trying to solve the linear MMSE estimation but also has a power constraint limiting the amount of data it is able to transmit per iteration.

In the proposed algorithm for utility based cross-layer collaboration, all of the microphones are included in the active set $\forall\omega_l$, $\mathbb{I}(\omega_l) = \emptyset$, so that there is initially full signal communication. The full signal is first used to estimate the correlation matrices and the MWF which is collected every STFT-frame t as defined in equation (1). After the full signal estimate has been collected the lowest utility is found, in a greedy fashion, over all of the signals available.

The utility given in equation (8) is observed over all frequency bins $U_k(\omega_l)$, where the minimum utility is given by

$$min\,\mathbf{u}_{\mathbb{A}} = min_{k\in\{2\ldots N\},l\in\{1\ldots L\}}U_k(\omega_l) \tag{18}$$

which does not include the reference microphone (k=1) in the calculation. The reference microphone is assumed to be attached to the fusion center so that the full signal is used for estimation and excluded from the calculation of the utility[1]. The correlation matrix and MWF are updated accordingly after every bin removal to ensure that they contain only statistics from the active set.

Once the power constraint has been satisfied the nodes transmit their signal with only the relevant signal components, deemed by the utility, at STFT intervals now represented by t_s to indicate that a subset of the full signal is being sent.

In order to ensure the adaptability of the system it is then beneficial to periodically re-estimate the utility and add microphones that contribute to the MMSE estimation using equation (17). Therefore the full signal needs to be transmitted periodically in order to re-evaluate the full signal characteristics. The nodes transmit their full signal at discrete time intervals t_f which is related to the STFT interval by a discrete factor n

$$t_f n = t_s. \tag{19}$$

During the full signal transmission the maximum utility is observed over all frequency bins much like equation (18) where

$$max\,\mathbf{u}_{\mathbb{I}} = max_{k\in\{2\ldots N\},l\in\{1\ldots L\}}U_k(\omega_l) \tag{20}$$

where again the reference microphone is excluded from the calculation. The highest utility value is then added to the current estimation and the MWF is re-calculated for every bin added. This process is re-iterated until one bin has been added for each microphone.

After bin addition, the bin deletion as outlined previously, is again applied to the signal estimate in order to eliminate the lowest contributing signal components. This deletion differs from the starting part of the algorithm as only one bin needs to be removed per microphone. This is due to the fact that since the power constraint has already been satisfied in the beginning of the algorithm and the bin addition only adds N bins over the constraint, only N bins need to be removed. The addition-deletion procedure is then repeated indefinitely throughout the signal estimation to guarantee the system is able to adapt to changes in the signal characteristics.

4.2 Filter estimate from delayed statistics

In order to adapt to the variation in signal characteristics in the proposed scenario, all bins are transmitted every t_f which keeps track of signal characteristics over a longer time frame than when a subset of the signal is transmitted at t_s. This time difference then introduces significant problems in the estimation of the correlation matrices and thus the optimal filter value.

The correlation matrices for full signal estimation can be tracked at the fusion center much like equation (6) where

$$\mathbf{R}_{\mathbf{yy}}[t_f] = \lambda\mathbf{R}_{\mathbf{yy}}[t_f - 1] + (1-\lambda)\mathbf{y}[t_f]\mathbf{y}[t_f]^H \tag{21}$$

which is updated with values from the active and inactive set. The active channels of the matrix are updated more frequently at intervals t_s given by

$$\mathbf{R}_{\mathbf{y}_{\mathbb{A}}\mathbf{y}_{\mathbb{A}}}[t_s] = \lambda\mathbf{R}_{\mathbf{y}_{\mathbb{A}}\mathbf{y}_{\mathbb{A}}}[t_s - 1] + (1-\lambda)\mathbf{y}_{\mathbb{A}}[t_s]\mathbf{y}_{\mathbb{A}}[t_s]^H. \tag{22}$$

For microphone deletion it was shown in section 3.1 that the utility is only calculated using the current active microphone statistics. The MWF can then be estimated from the statistics that have been transmitted every t_s.

[1] If the utility is also applied to the reference microphone bins would be removed from the desired speech source causing a loss to a portion of the spectrum and therefore impacting the signal estimate.

Algorithm 1: Proposed Utility Based cross-layer collaboration

initialization;
$\forall \omega_l, k : k \in \mathbb{A}(\omega_l), \mathbb{I}(\omega_l) = \emptyset$;
while *Power > Power Constraint* **do**
  Compute (10) $\forall k \in \mathbb{A}$;
  Find minimum $U_{k,l}$ (18);
  Move k from $\mathbb{A}$ to $\mathbb{I}$;
  update $\mathbf{R}_{\mathbf{y}_\mathbb{A}\mathbf{y}_\mathbb{A}}$, $\hat{\mathbf{w}}_\mathbb{A}$;
$t_s \leftarrow 0, t_f \leftarrow 0$;
repeat
  if *(t_s mod n) = 0* **then**
    all nodes transmit $\mathbf{y}[\omega_l, t]\ \forall \omega_l$;
    $\mathbf{R}_{\mathbf{yy}}[t_f + 1] = \lambda \mathbf{R}_{\mathbf{yy}}[t_f] + (1-\lambda)\mathbf{y}[t_f]\mathbf{y}[t_f]^H$;
    update $\mathbf{r}_{\mathbf{y}_\mathbb{A} d}$, $\mathbf{r}_{\mathbf{y}_\mathbb{I} d}$;
    update $\mathbf{R}_{\mathbf{y}_\mathbb{A}\mathbf{y}_\mathbb{I}}$ for (15,16);
    $\hat{\mathbf{w}} = \mathbf{R}^{-1}_{\mathbf{y}_\mathbb{A}\mathbf{y}_\mathbb{A}}[t_f + 1] r_{\mathbf{y}_\mathbb{A} d}$;
    $t_f \leftarrow t_f + 1$;
    for *1 to N* **do**
      Compute (17) $\forall k \in \mathbb{I}$;
      Find maximum $U_{k,l}$ using (20);
      Move k from $\mathbb{I}$ to $\mathbb{A}$;
      update $\mathbf{R}_{\mathbf{y}_\mathbb{A}\mathbf{y}_\mathbb{A}}$, $\hat{\mathbf{w}}_\mathbb{A}$;
    for *1 to N* **do**
      Compute (10) $\forall k \in \mathbb{A}$;
      Find minimum $U_{k,l}$ using (18);
      Move k from $\mathbb{A}$ to $\mathbb{I}$;
      update $\mathbf{R}_{\mathbf{y}_\mathbb{A}\mathbf{y}_\mathbb{A}}$, $\hat{\mathbf{w}}_\mathbb{A}$;
  $\hat{d}_\mathbb{A}[t_s] = \hat{\mathbf{w}}^H_\mathbb{A} \mathbf{y}_\mathbb{A}[t_s]$;
  $t_s \leftarrow t_s + 1$;
until *End of Data (EOD)*;

Adding microphones is a significantly different and more difficult problem. In section 3.2 the addition of microphones poses a problem as the correlation matrices depend on the active and inactive channels. In order to accurately represent the utility for microphone addition the fusion center must rely on data that has been collected from both the t_s and t_f intervals. The statistics in the full correlation matrices (21) and the $\mathbb{A}, \mathbb{I}$ submatrices must contain equivalent statistics across all microphones and frequencies in order to derive the optimal MWF.

One way to rectify this problem is to no longer update the filter and correlation matrices during t_s and to only apply a static filter to the received data until t_f is incremented. This allows for the utility to use values that have been collected over the same time interval.

Another method to reduce the mis-match introduced by using data from the t_f and t_s intervals is to switch back to the long term statistics every t_f interval. The MWF is updated every t_s interval using (22) and $\mathbf{r}_{\mathbf{y}_\mathbb{A} d}$. When t_f is incremented, the values in (21) corresponding to the active set then replace the values in (22). From this, $\mathbf{r}_{\mathbf{y}_\mathbb{A} d}$ is found and $\hat{\mathbf{w}}$ is re-calculated from the t_f statistics. The MWF now contains values from only the $t_f + 1$ and t_f intervals.

Unfortunately these methods directly hinder the adaptability of the MWF as it only stays valid if the signal characteristics change much slower then the transmission period for new data. This problem will be further addressed in section 5.

5. SIMULATIONS

In this section we simulated a scenario depicted in Figure 1. The simulated room environment contains one speech source ★, one white noise source ■, and a star-topology network with a fusion center • with a reference microphone, and 8 other nodes ○ each having one microphone. There is also added white noise to each sensor observation equal to 10% of the speech source power representative of thermal noise. The nodes, speech, and noise source are positioned at a height of 1.5 meters from the ground. A reverberation time of $T_{60} = 0.4$ s was used for the entire room.

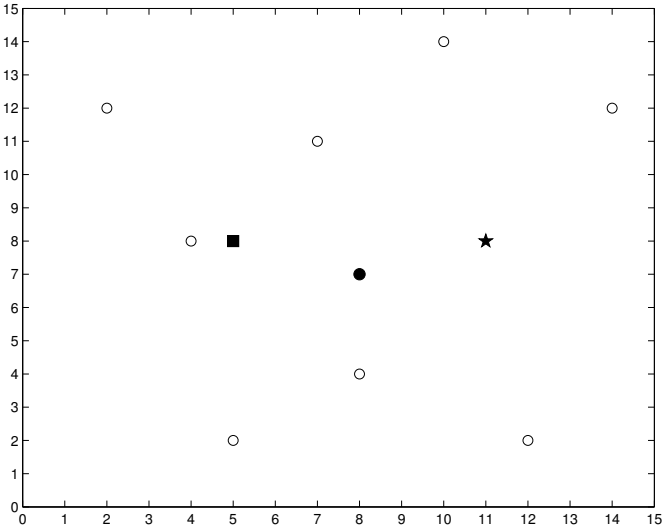

Figure 1: Simulated room scenario with a speech source ★, a white noise source ■, a fusion center with reference microphone •, and 8 single microphone nodes ○.

In order to reduce input-output delays on the system a weighted overlap add (WOLA) technique was used that was proposed in [5]. A DFT block length of 512 was used in order to evaluate signal statistics in the frequency domain. A forgetting factor of $\lambda = 0.997$ was used. The full signal was first collected for 5 seconds at a sampling frequency of $f_s = 8000$. The utility was then calculated from the collected data to determine which subset of frequency bins should be used during transmission. This subset was sent every iteration t_s and the full signal was transmitted after every 10 iterations, n = 10.

In order to assess the effect of sending a subset of the signal for estimation two rubrics were used. First a speech intelligibility weighted signal-to-noise ratio (IWSNR) [12, 13] was used in order to observe the effect of only using a subset of the signal in the estimation. While the IWSNR ratio is able to give characteristics pertaining to the speech and noise of the filtered signal little can be said about speech distortion.

The signal-to-error ratio (SER) was therefore used to monitor how much error is introduced by the compression. The SER is given by

$$SER = 10\log_{10}\left(\frac{\sum_{t_s} d[k]^2}{\sum_{t_s}(d[k] - \hat{d}[k])^2}\right) \quad (23)$$

where the denominator is the squared error between the signal and the filtered version of the signal.

Since only one microphone is considered per node the full signal collected from the nodes at the fusion center corresponds to $\mathbf{B}$ = NxL = $8x512 = 4068$ bins. The number of bins that are transmitted during every t_s interval is given by $\mathbf{B}_s$.

Figure 2a shows the IWSNR and the SER as a function of $\mathbf{B}_s$. Algorithm 1 was adjusted so that the statistics in (22) are replaced every t_f with the statistics from (21) as discussed in section 4.2. Reducing $\mathbf{B}_s$ corresponds to a gradual decrease in SNR as well as a drop in the SER.

Figure 2b explores the divergence of the IWSNR and SER from the full estimation procedure versus the amount of cycles between full and partial updates. In this scenario the filters are held constant during t_s as discussed in section 4.2. The SER and IWSNR for the full estimation are shown by the dashed lines as a function of increasing the ratio between t_s and t_f. There is very little drop in

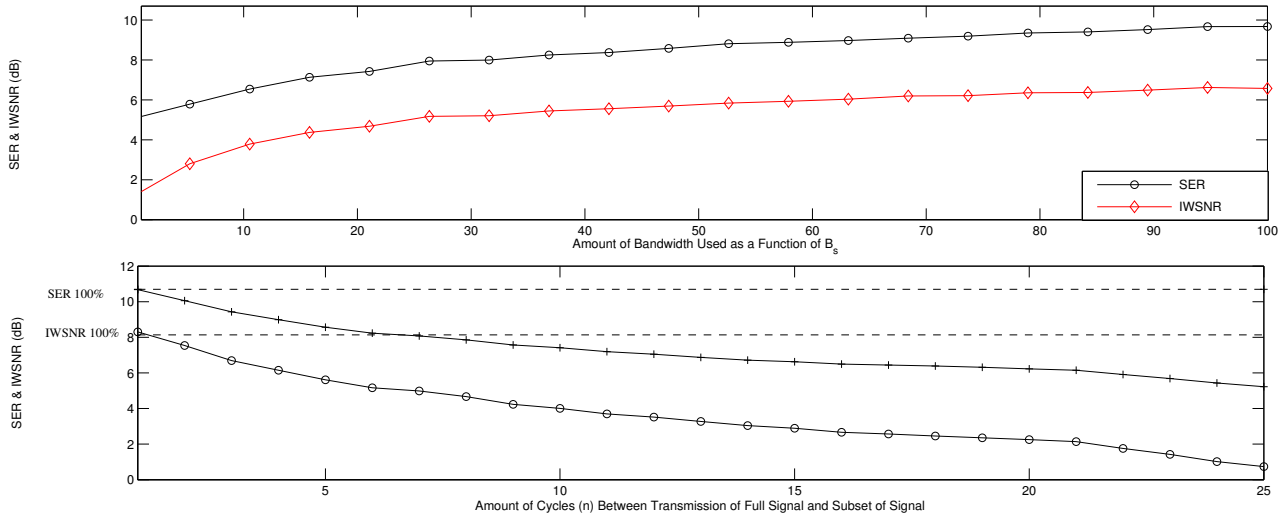

Figure 2: The top subplot shows both the SER and IWSNR as a function of $\mathbf{B_s}$. The bottom plot shows the difference between using a full signal and a subset of the signal while varying the time between full t_f and partial transmission t_s.

the SER going from 100% to 99% of the bandwidth when n is small $t_s \approx t_f$. The SER starts to diverge as t_f becomes much larger than t_s which can be attributed to the current signal estimation being based on an old signal characteristics at t_f. The more intermediate data that passes during n creates a larger difference in the matrices of (21,22).

6. CONCLUSION

A utility function has been used as a viable tool in order to best dictate which is the most important information available in a WASN. The network constraints of the system were able to be utilized to their full potential with little loss in signal quality. The utility also offers a more efficient way than placing a compression burden on the nodes in order to meet transmit power constraints, since utility measures are readily available from known variables in the estimation procedure. This method of reducing the transmission burden on the nodes also translates to energy saving of the entire WASN which ensures a longer lifetime of the system.

REFERENCES

[1] A. Goldsmith and S. Wicker, "Design challenges for energy-constrained ad hoc wireless networks," *Wireless Communications, IEEE*, vol. 9, no. 4, pp. 8 – 27, 2002.

[2] D. Estrin, L. Girod, G. Pottie, and M. Srivastava, "Instrumenting the world with wireless sensor networks," in *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on*, vol. 4, pp. 2033 –2036 vol.4, 2001.

[3] S. Doclo, M. Moonen, T. Van den Bogaert, and J. Wouters, "Reduced-bandwidth and distributed MWF-based noise reduction algorithms for binaural hearing aids," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 17, no. 1, pp. 38 –51, 2009.

[4] A. Bertrand and M. Moonen, "Robust distributed noise reduction in hearing aids with external acoustic sensor nodes," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, Article ID 530435, 14 pages, 2009.

[5] A. Bertrand, J. Callebaut, and M. Moonen, "Adaptive distributed noise reduction for speech enhancement in wireless acoustic sensor networks," in *Proc. of the International Workshop on Acoustic Echo and Noise Control (IWAENC)*, (Tel Aviv, Israel), August 2010.

[6] N. Kimura and S. Latifi, "A survey on data compression in wireless sensor networks," in *Information Technology: Coding and Computing, 2005. ITCC 2005. International Conference on*, vol. 2, pp. 8 – 13 Vol. 2, 2005.

[7] G. Gosztolya, D. Paczolay, and L. To andth, "Low-complexity audio compression methods for wireless sensors," in *Intelligent Systems and Informatics (SISY), 2010 8th International Symposium on*, pp. 77 –81, 2010.

[8] O. Roy and M. Vetterli, "Rate-constrained collaborative noise reduction for wireless hearing aids," *Signal Processing, IEEE Transactions on*, vol. 57, pp. 645 –657, feb. 2009.

[9] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 14, no. 4, pp. 1218 –1234, 2006.

[10] J. Benesty, M. M. Sondhi, and Y. A. Huang, *Springer Handbook of Speech Processing*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2007.

[11] A. Bertrand and M. Moonen, "Efficient sensor subset selection and link failure response for linear MMSE signal estimation in wireless sensor networks," in *Proc. of the European signal processing conference (EUSIPCO)*, (Aalborg - Denmark), pp. 1092–1096, August 2010.

[12] J. E. Greenberg, P. M. Peterson, and P. M. Zurek, "Intelligibility-weighted measures of speech-to-interference ratio and speech system performance," *J. Acoustic. Soc. Am.*, vol. 94, no. 5, pp. 3009–3010, Nov. 1993.

[13] Acoustical Society of America, "ANSI S3.5-1997 American National Standard Methods for calculation of the speech intelligibility index," June 1997.