DISTRIBUTED ESTIMATION AND EQUALIZATION OF ROOM ACOUSTICS IN A WIRELESS ACOUSTIC SENSOR NETWORK

Toon van Waterschoot and Marc Moonen

KU Leuven, Dept. of Electrical Engineering-ESAT / SCD-SISTA and IBBT Future Health Department
Kasteelpark Arenberg 10, 3001 Leuven, Belgium
{toon.vanwaterschoot,marc.moonen}@esat.kuleuven.be

ABSTRACT

In this paper, the use of a wireless acoustic sensor network (WASN) for the estimation and equalization of room acoustics is proposed as a flexible and promising alternative to the traditional wired implementations. We consider a multiple-point equalization problem based on a common-acoustical-pole (CAP) room model. Instead of collecting microphone signals in a central processing unit to compute the CAP model estimate in a centralized fashion, we deploy a large number of autonomous nodes with local sensing, processing, and communication capabilities to solve the CAP model estimation problem in a distributed manner. Even though the WASN nodes are restricted to exchange information with neighboring nodes only, the use of a distributed averaging algorithm results in a CAP model estimate with an accuracy and equalization performance comparable to a wired implementation.

Index Terms— Wireless acoustic sensor networks, room acoustics, equalization, distributed consensus averaging

1. INTRODUCTION

Room equalization is an important task in many acoustic signal processing applications, and is intended to flatten the frequency magnitude response of an acoustic enclosure (the "room"). In this way, the sound signals perceived at one or more listening positions in the room should ideally be close to the original ("dry") sound signal that one aims to play back. The room equalization problem is often approached as an inversion problem, in which an acoustic room model is estimated, inverted, and applied as a prefilter to the dry signal prior to playback [1]. Both the estimation and inversion of the room model are considered to be challenging tasks, due to the high-order and mixed-phase character of room acoustics [1]. Morever, in many applications, the aim is to achieve equalization at many –if not all– possible listening positions inside a room, a problem that is often referred to as multiple-point equalization [2].

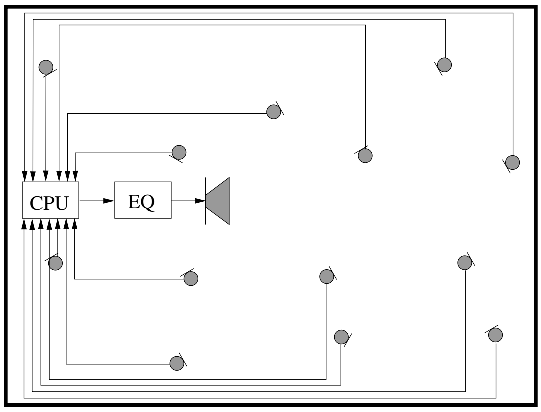


Fig. 1. Traditional implementation of multiple-point equalization system using wired microphones and a central processing unit (CPU).

One particularly interesting room model in this respect is the all-pole model [3]: it can be efficiently estimated using linear prediction, it can be straightforwardly inverted (resulting in an all-zero inverse filter), and it has been conjectured to be spatially invariant (i.e., independent of source and listener positions) inside a particular room [4]. The latter observation has motivated the use of the common-acoustical-pole (CAP) model as an alternative to the all-pole model [2]. The design of an equalization filter based on the CAP model consists of three steps [2]: (1) the estimation of the room impulse responses (RIRs) from the sound source position to a number of listening positions, (2) the estimation of the CAP model from the estimated RIRs, and (3) the calculation of an all-zero equalization filter by inverting the estimated CAP model. From a practical point of view, the first step of this procedure is the most challenging one, as it requires the deployment of a large number of microphones for collecting RIR measure-

T. van Waterschoot is a Postdoctoral Fellow of the Research Foundation Flanders (FWO–Vlaanderen). This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of KU 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 IBBT, and Research Project FWO nr. G.0600.08 "Signal processing and network design for wireless acoustic sensor networks". The scientific responsibility is assumed by its authors.

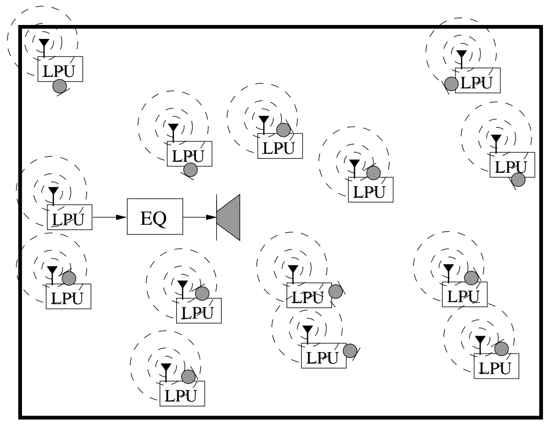


Fig. 2. Proposed implementation of multiple-point equalization system using a wireless acoustic sensor network (WASN) comprising wirelessly connected nodes with local processing units (LPUs).

ments at a wide range of positions in the room. Traditional implementations of a multiple-point equalization system rely on the use of wired microphones, such that signal measurements can be collected in a central processing unit (CPU) that computes the RIR and CAP model estimates, see Fig. 1.

In this paper, we propose a different implementation based on a wireless acoustic sensor network (WASN), as shown in Fig. 2. A WASN is a network of autonomous, battery-driven sensor nodes, each comprising a microphone, a local processing unit (LPU), and means for wireless communication with other sensor nodes and with the equalizer/loudspeaker node. In many applications, the WASN-based implementation is more appealing than the traditional one, due to its flexibility and ease of deployment. The installation, relocation, and addition of microphones is much easier in the WASN-based implementation, which makes this a particularly attractive solution when a multiple-point equalization system is to be appended to an existing sound reproduction system.

However, the equalization filter design in a WASN-based implementation cannot be executed using the traditional procedure outlined above, due to communication constraints inherent in the WASN. Indeed, since the WASN nodes are battery-driven, the bit rate and power at which these nodes can transmit data to other nodes is limited, and so the RIR and CAP model estimation procedure has to be fundamentally reorganized. Instead of transmitting all microphone signals to a CPU where the RIR and CAP model estimation is executed, we propose to distribute the RIR and CAP model estimation tasks over the LPUs available in the WASN while allowing WASN nodes to transmit a miminal amount of data only to neighboring nodes. We will show that the latter approach is feasible by formulating the CAP model estimation as a consensus problem, and employing a distributed averaging algorithm as proposed in [5].

The paper is organized as follows. In Section 2, we formulate the multiple-point equalization problem and define the WASN and its topology. Section 3 deals with the distributed estimation of the CAP model, while Section 4 contains results of Monte Carlo simulations and their discussion. Finally, Section 5 concludes the paper.

2. PROBLEM STATEMENT

We assume M microphones are deployed at positions $\mathbf{r}_m,\ m = 1,\ldots,M$ in a room where a sound signal $x(t)$ is played back using a loudspeaker at position $\mathbf{r}_s$. The resulting microphone signals are given by

$$y_m(t) = H(q,\mathbf{r}_s,\mathbf{r}_m)x(t) + v_m(t),\ m = 1,\ldots,M \quad (1)$$

where $H(q,\mathbf{r}_s,\mathbf{r}_m),\ m = 1,\ldots,M$ denote the length-L room impulse responses (RIRs) from the loudspeaker to the microphones, q is the time shift operator (i.e., $q^{-k}x(t) = x(t-k)$), and $v_m(t),\ m = 1,\ldots,M$ represents measurement noise at the microphones. The idea of multiple-point equalization is to play back a prefiltered signal $u(t) = F(q)x(t)$ instead of the original signal $x(t)$, and to calculate the equalization filter $F(q)$ such that the resulting microphone signals $y_m(t),\ m = 1,\ldots,M$ are perceived as closely as possible to the original signal $x(t)$. A convenient way of designing the equalization filter results from representing the RIRs as follows [2],[3],

$$H(q,\mathbf{r}_s,\mathbf{r}_m) = \frac{\tilde{H}(q,\mathbf{r}_s,\mathbf{r}_m)}{A(q)} \quad (2)$$

where $A^{-1}(q) = (1 + a_1q^{-1} + \ldots + a_Pq^{-P})^{-1}$ is the CAP model of order P, and $\tilde{H}(q,\mathbf{r}_s,\mathbf{r}_m)$ denote the residual RIRs. The equalization filter is then chosen as the inverse of the estimated CAP model, i.e., $F(q) = \hat{A}(q)$. The choice of equalizing only the CAP model $A^{-1}(q)$ and not the residual RIR component $\tilde{H}(q,\mathbf{r}_s,\mathbf{r}_m)$ is justified by the assumption that room resonances contribute most to the perceived difference between the loudspeaker and microphone signals.

The estimation of the CAP model $A^{-1}(q)$ is usually based on available RIR estimates

$$\hat{H}(q,\mathbf{r}_s,\mathbf{r}_m) = H(q,\mathbf{r}_s,\mathbf{r}_m) + E(q,\mathbf{r}_s,\mathbf{r}_m) \quad (3)$$

where the RIR estimation error $E(q,\mathbf{r}_s,\mathbf{r}_m)$ results from measurement noise at the microphones. While the traditional implementation in Fig. 1 allows for an on-line RIR estimation (since $x(t)$ is available in the CPU), the WASN-based implementation in Fig. 2 requires a training phase during which the RIR $H(q,\mathbf{r}_s,\mathbf{r}_m)$ is estimated in the LPU of the mth node, based on a known training signal $x(t)$. In both cases, however, the measurement noise and hence the RIR estimation error has the same variance.

The topology of the WASN is determined by the sensor node positions $\mathbf{r}_m,\ m = 1,\ldots,M$, the equalizer/loudspeaker node position $\mathbf{r}_s$ and the assumed communication model. Here, we adopt a simple communication model, where error-free communication between sensor nodes k and l is possible if $(\mathbf{r}_k - \mathbf{r}_l)^T(\mathbf{r}_k - \mathbf{r}_l) \leq \rho^2$ while no communication is possible otherwise, and likewise for the communication between the sensor nodes and the equalizer/loudspeaker node. In other words, the WASN nodes only communicate with neighboring nodes, where the neighborhood is defined by the communication range ρ. The WASN topology can hence be represented by the symmetric $M \times M$ sensor connectivity matrix $\mathbf{C}$, defined as

$$[\mathbf{C}]_{kl} = \begin{cases} 1, & \text{if } (\mathbf{r}_k - \mathbf{r}_l)^T(\mathbf{r}_k - \mathbf{r}_l) \leq \rho^2 \quad \text{(4a)} \\ 0, & \text{if } (\mathbf{r}_k - \mathbf{r}_l)^T(\mathbf{r}_k - \mathbf{r}_l) > \rho^2 \quad \text{(4b)} \end{cases}$$

and the neighborhood $\mathcal{N}_s = \{m | (\mathbf{r}_m - \mathbf{r}_s)^T(\mathbf{r}_m - \mathbf{r}_s) \leq \rho^2\}$ of the equalizer/loudspeaker node.

3. CAP MODEL ESTIMATION

3.1. Traditional implementation: least squares (LS) and centralized averaging (CAV)

In a traditional implementation, all RIR estimates $\hat{H}(q, \mathbf{r}_s, \mathbf{r}_m)$, $m = 1, \ldots, M$ are available in the CPU, and the least-squares (LS) estimate of the CAP model parameter vector $\mathbf{a} = [a_1, \ldots, a_P]^T$ can be computed by a linear prediction of the concatenated and zero-padded estimated RIR parameter vectors, i.e.,

$$\hat{\mathbf{a}}_{\text{LS}} = \left(\sum_{m=1}^{M} \hat{\mathbf{H}}_m^T \hat{\mathbf{H}}_m\right)^{-1} \left(\sum_{m=1}^{M} \hat{\mathbf{H}}_m^T \hat{\mathbf{h}}_m\right) \tag{5}$$

(see [2] for a definition of $\hat{\mathbf{H}}_m$ and $\hat{\mathbf{h}}_m$). However, an interesting observation in [4] is that the CAP model parameter vector estimate in (5) is closely approximated by a centralized averaging (CAV) of the local (i.e., node-specific) all-pole model parameter vector estimates resulting from a linear prediction of the local RIR parameter vectors, i.e.,

$$\hat{\mathbf{a}}_{\text{LS}} \approx \hat{\mathbf{a}}_{\text{CAV}} = \sum_{m=1}^{M} \hat{\mathbf{a}}_{m,\text{LS}} = \sum_{m=1}^{M} \left[\left(\hat{\mathbf{H}}_m^T \hat{\mathbf{H}}_m\right)^{-1} \hat{\mathbf{H}}_m^T \hat{\mathbf{h}}_m\right]. \tag{6}$$

3.2. WASN-based implementation I: localized averaging (LAV)

In a WASN-based implementation, the estimates in (5) and (6) can generally not be calculated since none of the WASN nodes has access to all local RIR or all-pole model estimates. Moreover, communicating local RIR estimates between neighboring nodes should be avoided due to the requirement of low bit rates (which conflicts with the typically high RIR lengths). A straightforward yet suboptimal approach to estimate the CAP model parameter vector then consists in collecting and averaging the local LS all-pole model parameter vector estimates from the sensor nodes in the neighborhood $\mathcal{N}_s$ of the equalizer/loudspeaker node, i.e.,

$$\hat{\mathbf{a}}_{\text{LAV}} = \sum_{m \in \mathcal{N}_s} \hat{\mathbf{a}}_{m,\text{LS}}. \tag{7}$$

In this case, the sensor nodes outside $\mathcal{N}_s$ do not contribute to the CAP model estimate, hence this approach is denoted as localized averaging (LAV).

3.3. WASN-based implementation II: distributed averaging (DAV)

Alternatively, the estimates in (5) and (6) can be cast into a consensus optimization framework. The LS estimate in (5) can hence be approximated by solving M local LS problems including a consensus constraint, i.e.,

$$\{\hat{\mathbf{a}}_{m,\text{DLS}}\}_{m=1}^{M} = \arg\min_{\mathbf{a}_m} \sum_{m=1}^{M} \|\hat{\mathbf{H}}_m \mathbf{a}_m - \mathbf{h}_m\|_2^2 \quad \text{s.t. } \mathbf{a}_m = \hat{\mathbf{a}}_{\text{DLS}}. \tag{8}$$

This distributed LS problem, where $\hat{\mathbf{a}}_{\text{DLS}}$ denotes the consensus CAP model parameter vector estimate, can be iteratively solved using the alternating direction method of multipliers (ADMoM) [6].

A simpler and equally accurate approach is to compute the average of all local all-pole model parameter vector estimates using a distributed averaging (DAV), which only requires local communication among neighboring nodes. A fast distributed linear averaging (FDLA) algorithm [5] is defined by the iteration

$$\left[\hat{\mathbf{a}}_{1,\text{FDLA}}^{(k)} \ \ldots \ \hat{\mathbf{a}}_{M,\text{FDLA}}^{(k)}\right] = \left[\hat{\mathbf{a}}_{1,\text{FDLA}}^{(k-1)} \ \ldots \ \hat{\mathbf{a}}_{M,\text{FDLA}}^{(k-1)}\right] \mathbf{W}_{\text{FDLA}}, \quad k = 1, \ldots, k_{\max} \tag{9}$$

in which the initialization correponds to the local LS all-pole model parameter vector estimates,

$$\hat{\mathbf{a}}_{m,\text{FDLA}}^{(0)} = \hat{\mathbf{a}}_{m,\text{LS}}, \ m = 1, \ldots, M \tag{10}$$

and the optimal (symmetric) weighting matrix is calculated by solving the following convex optimization problem [5]

$$\mathbf{W}_{\text{FDLA}} = \arg\min_{\mathbf{W}} \|\mathbf{W} - \mathbf{1}\mathbf{1}^T/M\|_2 \tag{11}$$

$$\text{s.t. } \mathbf{W} \in \mathcal{S}(\mathbf{C}), \ \mathbf{1}^T\mathbf{W} = \mathbf{1}^T, \ \mathbf{W}\mathbf{1} = \mathbf{1} \tag{12}$$

where $\mathbf{1}$ is a length-M column vector with all ones and $\mathcal{S}(\mathbf{C})$ denotes the class of matrices having the same sparsity pattern as the sensor connectivity matrix $\mathbf{C}$. After the final iteration of the algorithm in (9), the CAP model parameter vector estimate is calculated as

$$\hat{\mathbf{a}}_{\text{DAV}} = \sum_{m \in \mathcal{N}_s} \hat{\mathbf{a}}_{m,\text{FDLA}}^{(k_{\max})}. \tag{13}$$

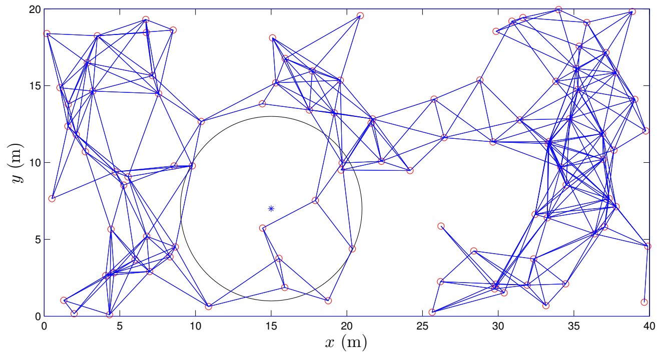


Fig. 3. 2-D projection on $\{x, y\}$ plane of 3-D WASN topology using $J = 100$ sensor nodes (o) and communication range $\rho = 6$ m. Blue lines denote communication links between sensor nodes; black circle indicates communication range of equalizer/loudspeaker node ($*$).

4. SIMULATION RESULTS

The evaluation of the multiple-point equalization implementations discussed in Section 3 is based on the average performance over $N = 100$ Monte Carlo trials of a WASN comprising $M = 100$ sensor nodes deployed at random positions. Since a database of $MN = 10000$ RIR measurements is currently not availabe, we resort to a simulated acoustic environment. A reverberant $40 \times 20 \times 10$ m shoe-box shaped room is simulated based on a CAP model $A^{-1}(q)$ of order $P = 24$ calculated by pole placement, and residual RIRs $\tilde{H}(q, \mathbf{r}_s, \mathbf{r}_m)$, $m = 1, \ldots, M$ with $\mathbf{r}_s^T = [15, 7, 7]$ m generated using the image source method [7]. The RIRs $H(q, \mathbf{r}_s, \mathbf{r}_m)$ are then obtained by truncating the impulse responses resulting from the CAP model in (2) to a length of $L = 2000$, corresponding to 0.25 s when sampling at 8 kHz.

As explained in Section 2, the measurement noise at the microphones leads to a RIR estimation error $E(q, \mathbf{r}_s, \mathbf{r}_m)$. We simulate this effect by directly adding Gaussian white noise to the estimated local all-pole model coefficients, which is equivalent to using a spectrally flat training signal $x(t)$ for estimating the local all-pole models, and assuming Gaussian white measurement noise. The resulting RIR estimation accuracy, defined as $10\log_{10} \|\mathbf{a}\|_2^2 / \|\mathbf{a} - \hat{\mathbf{a}}_{m,\text{LS}}\|_2^2$, is fixed to an average value of 10 dB. The communication range of the WASN nodes is set to $\rho = 6$ m, which results in the (projected) topology shown in Fig. 3 for one particular realization of the sensor node positions. The number of iterations used in the FDLA algorithm (9) is set to $k_{\max} = 100$, which allows the DAV estimate (13) to converge to a value close to the CAV estimate (6). We should note, however, that we have observed the DAV algorithm to outperform the LAV algorithm for all values $k_{\max} \geq 1$.

We compare the resulting CAV, LAV, and DAV estimates defined in (6), (7), and (13) in terms of two equalization performance measures and one estimation performance measure.

The equalization performance is measured by assessing the spectral flatness of the residual RIRs, using the spectral flatness measure (SFM) [8, Ch. 6]

$$\sum_{n=1}^{N} \sum_{m=1}^{M} \frac{10}{MN} \log_{10} \frac{\exp\left[\frac{1}{L}\sum_{k=0}^{L-1} \ln \left|\tilde{H}^{(n)}\left(e^{j\frac{2\pi k}{L}}, \mathbf{r}_s, \mathbf{r}_m\right)\right|^2\right]}{\frac{1}{L}\sum_{k=0}^{L-1} \left|\tilde{H}^{(n)}\left(e^{j\frac{2\pi k}{L}}, \mathbf{r}_s, \mathbf{r}_m\right)\right|^2} \tag{14}$$

and the standard deviation (STD) [2]

$$\frac{1}{MN}\sum_{n=1}^{N}\sum_{m=1}^{M}\left(\frac{1}{L}\sum_{k=0}^{L-1}\left(10\log_{10}\left|\tilde{H}^{(n)}\left(e^{j\frac{2\pi k}{L}}, \mathbf{r}_s, \mathbf{r}_m\right)\right|^2 - \frac{1}{L}\sum_{l=0}^{L-1} 10\log_{10}\left|\tilde{H}^{(n)}\left(e^{j\frac{2\pi l}{L}}, \mathbf{r}_s, \mathbf{r}_m\right)\right|^2\right)^2\right)^{\frac{1}{2}}. \tag{15}$$

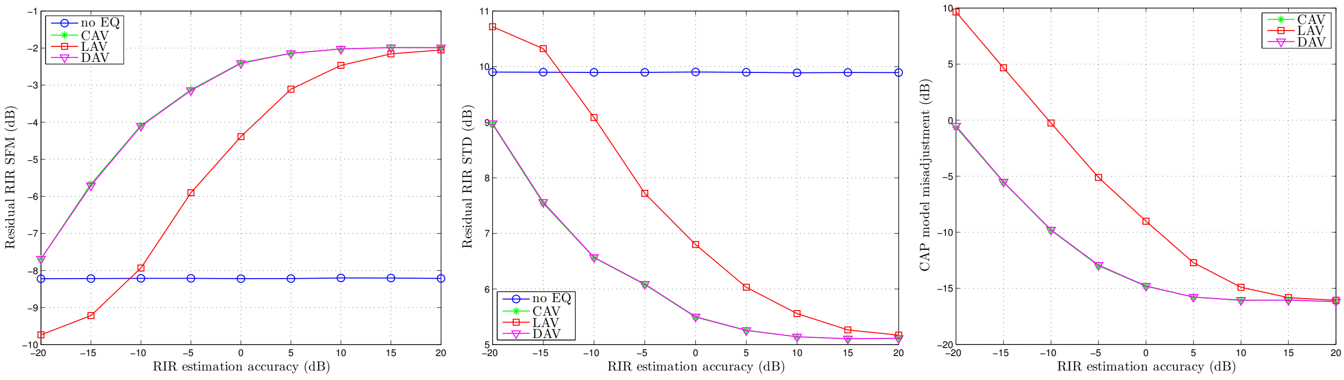
The estimation performance is measured by the misadjustment of the CAP model parameter vector,

$$10\log_{10}\left(\frac{1}{N}\sum_{n=1}^{N} \frac{\|\mathbf{a} - \hat{\mathbf{a}}^{(n)}_{\text{(C)(L)(D)AV}}\|_2^2}{\|\mathbf{a}\|_2^2}\right). \tag{16}$$

The resulting performance measures are plotted versus the RIR estimation accuracy and the WASN communication range in Figs. 4 and 5, respectively. For a communication range of $\rho = 6$ m, the DAV performance is equal to the CAV performance, regardless of the RIR estimation accuracy (i.e., the DAV and CAV curves overlap in Fig. 4). The LAV performance, on the other hand, is consistently worse and approaches the CAV performance only for high RIR estimation accuracy. From Fig. 5, it can be seen that the DAV performance breaks down for communication range values $\rho \leq 4$ m, which is explained by the fact that in this case the WASN does not correspond to a connected graph (which is a fundamental condition for convergence of the FDLA algorithm [5]). However, even for $\rho \leq 4$ m, the DAV performance is consistently better than the LAV performance.

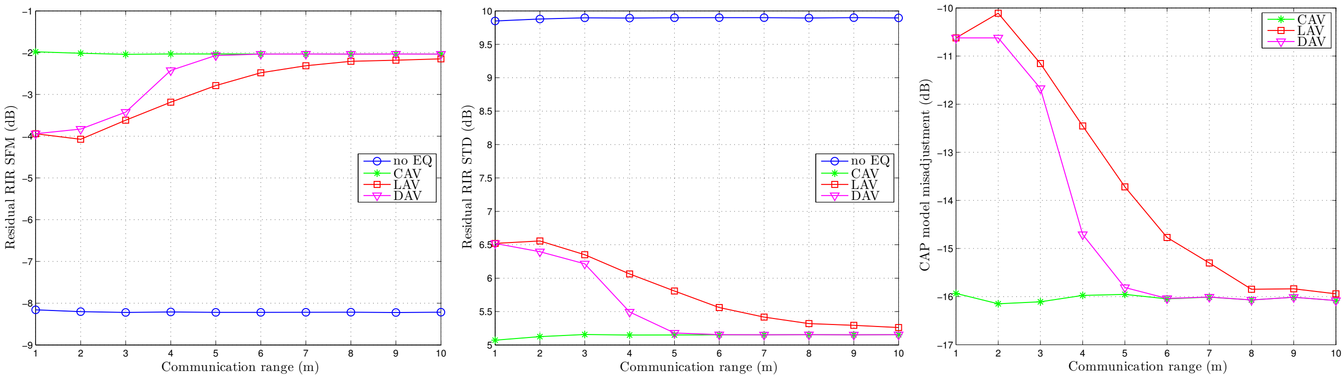
5. CONCLUSION AND FUTURE WORK

We have proposed a fundamentally new implementation for a multiple-point equalization system based on a CAP room model. By replacing wired microphones with a WASN, and distributing the processing effort, an easily deployed and flexible equalization system is obtained. Two different approaches for estimating the CAP model in a WASN-based implementation have been put forward: a localized averaging algorithm which only relies on information provided by sensor nodes in the neighborhood of the equalizer/loudspeaker node, and a distributed averaging algorithm in which information from all sensor nodes is used. Simulation results have shown that the distributed averaging approach results



(a) (b) (c)

Fig. 4. Performance vs. RIR estimation accuracy: (a) Residual RIR SFM, (b) Residual RIR STD, (c) CAP model misadjustment.



(a) (b) (c)

Fig. 5. Performance vs. communication range: (a) Residual RIR SFM, (b) Residual RIR STD, (c) CAP model misadjustment.

in a consistently better performance than the centralized averaging approach. Moreover, if the communication range is sufficiently large such that the WASN corresponds to a connected graph, then the distributed averaging approach results in a performance similar to the centralized averaging approach used in a traditional wired implementation.

Two research challenges have been postponed to future work. First, the estimation and equalization performance of the WASN-based implementation should be validated using measured rather than simulated RIRs. Second, a more realistic communication model should be adopted, which also takes into account quantization effects and channel noise in the wireless communication between the WASN nodes.

6. REFERENCES

[1] J. N. Mourjopoulos, "Digital equalization of room acoustics," *J. Audio Eng. Soc.*, vol. 42, no. 11, pp. 884–900, Nov. 1994.

[2] Y. Haneda, S. Makino, and Y. Kaneda, "Multiple-point equalization of room transfer functions by using common acoustical poles," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 4, pp. 325–333, July 1997.

[3] J. Mourjopoulos and M. A. Paraskevas, "Pole and zero modeling of room transfer functions," *J. Sound Vib.*, vol. 146, no. 2, pp. 281–302, Apr. 1991.

[4] Y. Haneda, S. Makino, and Y. Kaneda, "Common acoustical pole and zero modeling of room transfer functions," *IEEE Trans. Speech Audio Process.*, vol. 2, no. 2, pp. 320–328, Apr. 1994.

[5] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Control Lett.*, vol. 53, no. 1, pp. 65–78, Sept. 2004.

[6] I. D. Schizas, A. Ribeiro, and G. B. Giannakis, "Consensus in ad hoc WSNs with noisy links – Part I: Distributed estimation of deterministic signals," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 350–364, Jan. 2008.

[7] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[8] J. D. Markel and A. H. Gray, Jr., *Linear prediction of speech*, Springer-Verlag, New York, 1976.