Online estimation of time-variant microphone utility in wireless acoustic sensor networks using single-channel signal features

Michael Günther, Andreas Brendel, Walter Kellermann
Multimedia Communications and Signal Processing, Friedrich-Alexander Universität Erlangen-Nürnberg (FAU),
{michael.guenther,andreas.brendel,walter.kellermann}@fau.de

***Abstract*—Rating the usefulness of individual microphones for subsequent signal processing is a key problem in Wireless Acoustic Sensor Networks (WASNs). This contribution expands a general-purpose approach for microphone ranking and selection in WASNs to facilitate online selection in time-variant scenarios. A Kalman Filter (KF) is employed to robustly estimate the time-varying inter-channel correlation of single-channel signal features. The information contained in multiple features is efficiently combined by recursively updating the dominant singular vector of each channel-wise matrix of feature correlation coefficients to obtain the similarity of the recorded signals while accounting for all features simultaneously. Capturing the resulting pair-wise similarity of microphone channels by a graph structure, the individual microphones' utility is obtained as the Fiedler vector, which is an eigenvector of the graph Laplacian. A method to resolve the inherent sign ambiguity of the Fiedler vector using the entropy of the observed microphone signals is proposed. Experiments using synthesized and recorded data demonstrate the efficacy of the proposed approach.**

***Index Terms*—channel selection, Fiedler vector, graph Laplacian, microphone utility, wireless acoustic sensor network**

I. Introduction

Multichannel acoustic signal processing algorithms, e. g., for spatial filtering [1] or source localization [2], rely on spatially distributed microphones offering different views on the acoustic scene. While this is especially true for Wireless Acoustic Sensor Networks (WASNs) with their potentially very large inter-microphone distances, there, the available microphone signals are generally not equally useful for subsequent signal processing applications. Their usefulness depends on a variety of factors, e. g., proximity to acoustic sources, transducer directivity and orientation, occlusion of sensors, reverberation and additive sensor noise [3]. Furthermore, a microphone's usefulness is application-specific in the sense that some signals can be useful to specific signal processing algorithms, e. g., as a noise reference for a Generalized Sidelobe Canceller (GSC) [4], while being useless for other algorithms.

We assume the observed sound field comprises a single coherent sound field component evoked by the Source of Interest (SoI), and diffuse or incoherent components originating from typical signal degradations, like additive sensor noise, reverberation and spatially extended interferers [5]. With this in mind, correlation between the observed signals and other quantities derived from it appear as a natural choice to quantify how useful microphone signals are for coherent signal processing. Prior published work includes correlation-based selection schemes for a fixed number of microphones for beamforming based on the MultiChannel Correlation Coefficient (MCCC) [6] and microphone utility measures specifically for Minimum Mean Square Error (MMSE) signal extraction [7]. In WASNs, the need to collect the microphone signals at a central unit to be able to compute the utility measure constitutes a significant drawback for both aforementioned methods (a distributed version of [7] has been proposed in [8]). A heuristic selection of the single best channel in terms of Word Error Rate (WER) in Automatic Speech Recognition (ASR) based on signal features has been proposed in [9], but requires ASR systems to be available for utility rating. In contrast, a general-purpose method to estimate microphone utility that exploits the correlation of low-dimensional single-channel signal features as a compressed representation of the signal waveform for both channel selection and economizing data rate in WASNs was proposed in [10], [11]. However, not transmitting the waveforms precludes usage of cross-channel features, like coherence. As the approach does not optimize cost functions and performance measures of subsequent applications, it remains applicable for a broad class of algorithms.

In this paper, we expand the work in [10], [11] to account for time-variant microphone utility caused, e. g., by moving acoustic sources and sudden changes of the acoustic environment typical for smart home scenarios. We assume that the SoI is the only common signal component that evokes similar feature values in all microphone signals and that the SoI dominates the noise components to avoid similar noise statistics dominating the feature correlation. Thus, the relative contribution of the SoI to each microphone signal can be estimated by the dominant singular vector of each channel-wise matrix of feature correlation coefficients [10] that best explains the observed feature covariance, termed similarity vector for brevity. We employ a Kalman Filter (KF) for robust online estimation of the feature correlation coefficients. To facilitate efficient tracking of the time-variant microphone utility, we provide a recursive update rule for the dominant singular vectors that avoids recomputing costly Singular Value Decompositions (SVDs) in every time step. Finally, we propose an improved scheme to resolve the sign ambiguity of the Fiedler vector based on the differential entropy of the observed signals.

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 282835863 – within the Research Unit FOR2457 "Acoustic Sensor Networks".

Algorithm 1 Proposed update for microphone utility at time k

Input: $\mathbf{a}^{(f)}[k]$ (instantaneous feature vector)
Input: $\boldsymbol{\mu}^{(f)}[k-1], \quad \forall f \in \{1,\dots,F\}$ (previous KF state vectors)
Input: $\mathbf{g}_p[k-1], \quad \forall p \in \{1,\dots,P\}$ (previous similarity vectors)
Output: $\mathbf{u}[k]$ (Fiedler vector)
  for features $f = 1$ to F **do**
    $\boldsymbol{\mu}^{(f)}[k] \leftarrow$ updated KF state vector, see (9) and (11) to (13)
    $b_{p,p'} \leftarrow$ updated feature correlation coefficients, see (14) and (15)
  end for
  for microphones $p = 1$ to P **do**
    $\mathbf{g}_p[k] \leftarrow$ updated similarity vector, see (16) to (18)
  end for
  $\mathbf{R}[k] \leftarrow$ concatenated normalized $\mathbf{g}_p[k], \ \forall p \in \{1,\dots,P\}$, see (19)
  $\mathbf{W}[k] \leftarrow$ symmetric adjacency matrix, see (20)
  $\mathbf{L}[k] \leftarrow$ random-walk graph Laplacian, see [11]
  $\mathbf{u}[k] \leftarrow$ Fiedler vector of $\mathbf{L}$, sign correction using entropy (21)

II. Online Estimation of Microphone Utility

In this section, we expand the method in [11] to facilitate online estimation of the microphone utility and specifically highlight the advances relative to [11]. The proposed method comprises three parts, described in Sections II-A to II-C:

1) computing the correlation coefficients of individual signal features across microphones
2) computing the pair-wise channel similarity w. r. t. several features simultaneously, based on the similarity vector, i. e., the dominant singular vector of the matrix of feature correlation coefficients in each microphone channel
3) constructing a graph structure to model the pair-wise channel similarity and computing the Fiedler vector of its random-walk graph Laplacian

A summary of the proposed method is given in Algorithm 1.

A. Feature Correlation Coefficients

Let $a_p^{(f)}[k]$ denote the value of feature $f \in \{1,\dots,F\}$ observed at microphone $p \in \{1,\dots,P\}$ for signal block k and define the instantaneous feature vector

$$\mathbf{a}^{(f)}[k] = \begin{bmatrix} a_1^{(f)}[k] & \cdots & a_P^{(f)}[k] \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^P. \tag{1}$$

Possible features include signal energy, statistical moments of the corresponding waveform or spectrum, and many more [12], [13]. In [11], the feature correlation coefficients $\tilde{b}_{p,p'}^{(f)}$ of microphones p, p' w. r. t. feature f are computed offline using knowledge of the feature values in all signal blocks. Instead, we cast the recursive estimation of the covariance matrix of $\mathbf{a}^{(f)}[k]$ as a KF and obtain the feature correlation coefficients $\tilde{b}_{p,p'}^{(f)}[k]$ by subsequent normalization of the estimate as detailed in the following. Note that each feature requires a separate KF. To this end, we denote the outer product of the instantaneous feature vector $\mathbf{a}^{(f)}[k]$ after subtracting the sample average by

$$\mathbf{A}^{(f)}[k] = \left(\mathbf{a}^{(f)}[k] - \overline{\mathbf{a}}^{(f)}[k]\right)\left(\mathbf{a}^{(f)}[k] - \overline{\mathbf{a}}^{(f)}[k]\right)^{\mathrm{T}}, \tag{2}$$

where $\overline{\mathbf{a}}^{(f)}[k]$ is the recursive sample average of $\mathbf{a}^{(f)}[k]$ with temporal averaging constant of λ. Then, the estimated covariance matrix of $\mathbf{a}^{(f)}[k]$ corresponding to feature f is

$$\mathbf{B}^{(f)}[k] = \begin{bmatrix} b_{1,1}^{(f)}[k] & \cdots & b_{1,P}^{(f)}[k] \\ \vdots & & \vdots \\ b_{P,1}^{(f)}[k] & \cdots & b_{P,P}^{(f)}[k] \end{bmatrix} = \hat{\mathcal{E}}\left(\mathbf{A}^{(f)}[k]\right), \tag{3}$$

where $\hat{\mathcal{E}}$ is an operator approximating statistical expectation by temporal averaging. Let the latent KF state vector $\mathbf{z}^{(f)}[k] \sim \mathcal{N}\left(\mathbf{z}^{(f)}[k] \middle| \boldsymbol{\mu}^{(f)}[k], \mathbf{P}^{(f)}[k]\right)$ be normally distributed with mean $\boldsymbol{\mu}^{(f)}[k]$ and covariance matrix $\mathbf{P}^{(f)}[k]$. We employ its mean vector $\boldsymbol{\mu}^{(f)}[k]$ to model $\mathbf{B}^{(f)}[k]$ as defined in (3). Due to the symmetry of $\mathbf{B}^{(f)}[k]$, it is sufficient to only estimate its non-redundant elements, such that $\boldsymbol{\mu}^{(f)}[k]$ comprises only the $Q = \frac{P(P+1)}{2}$ elements of the lower triangular portion of $\mathbf{B}^{(f)}[k]$ including the diagonal. Using the *half-vectorization* operator vech [14], the mean vector $\boldsymbol{\mu}^{(f)}[k]$ of the latent vector $\mathbf{z}^{(f)}[k]$ can be compactly written as

$$\boldsymbol{\mu}^{(f)}[k] = \operatorname{vech}\left(\mathbf{B}^{(f)}[k]\right) \in \mathbb{R}^Q. \tag{4}$$

Similarly, the KF observation vector $\mathbf{x}^{(f)}[k]$ consists of the half-vectorized outer product of the instantaneous feature vector after subtracting the sample average (see (2)), i. e.,

$$\mathbf{x}^{(f)}[k] = \operatorname{vech}\left(\mathbf{A}^{(f)}[k]\right). \tag{5}$$

Completing the KF model, the Q-variate Gaussian transition and emission distributions are

$$p\left(\mathbf{z}^{(f)}[k] \middle| \mathbf{z}^{(f)}[k-1]\right) = \mathcal{N}\left(\mathbf{z}^{(f)}[k] \middle| \mathbf{z}^{(f)}[k-1], \mathbf{S}\right), \tag{6}$$

$$p\left(\mathbf{x}^{(f)}[k] \middle| \mathbf{z}^{(f)}[k]\right) = \mathcal{N}\left(\mathbf{x}^{(f)}[k] \middle| \mathbf{z}^{(f)}[k], \mathbf{T}[k]\right), \tag{7}$$

respectively. The process noise covariance matrix is chosen as

$$\mathbf{S} = \alpha_1 \mathbf{I}_Q, \tag{8}$$

where α_1 is a positive scaling parameter and $\mathbf{I}_Q$ denotes the $Q \times Q$ identity matrix. We consider the observed features to be more reliable when the recorded signal has high energy, thus the observation noise covariance is chosen[1] inversely proportional to the geometric mean of the associated channel's signal energy, i. e.,

$$\mathbf{T}[k] = \alpha_2 \operatorname{Diag}\left(\operatorname{vech}\left(\mathbf{E}[k]\right)\right) \in \mathbb{R}^{Q \times Q}, \tag{9}$$

where α_2 is a positive scaling parameter and the Diag operator yields a diagonal matrix containing the elements of its argument on the main diagonal. The matrix $\mathbf{E}[k] \in \mathbb{R}^{P \times P}$ is defined elementwise

$$\left[\mathbf{E}[k]\right]_{pp'} = \left(\sqrt{e_p[k] \cdot e_{p'}[k]} + \epsilon\right)^{-1}, \tag{10}$$

where ϵ is a small positive constant to ensure invertibility during speech pauses and $e_p[k]$ denotes the energy of the p-th microphone's signal at time k. With the Kalman gain matrix

$$\mathbf{K}^{(f)}[k] = \left(\mathbf{P}^{(f)}[k] + \mathbf{S}\right)\left(\mathbf{P}^{(f)}[k] + \mathbf{S} + \mathbf{T}[k]\right)^{-1}, \tag{11}$$

the update equations for $\boldsymbol{\mu}^{(f)}[k]$ and $\mathbf{P}^{(f)}[k]$ read

$$\boldsymbol{\mu}^{(f)}[k+1] = \boldsymbol{\mu}^{(f)}[k] + \mathbf{K}^{(f)}[k]\left(\mathbf{x}^{(f)}[k] - \boldsymbol{\mu}^{(f)}[k]\right), \tag{12}$$

$$\mathbf{P}^{(f)}[k+1] = \left(\mathbf{P}^{(f)}[k] + \mathbf{S}\right)\left(\mathbf{I}_Q - \mathbf{K}^{(f)}[k]\right), \tag{13}$$

respectively. Finally, the feature correlation coefficients $\tilde{b}_{p,p'}^{(f)}[k], \ \forall p, p' \in \{1,\dots,P\}$ are computed by

$$\tilde{b}_{p,p'}^{(f)}[k] = \frac{b_{p,p'}^{(f)}[k]}{\sqrt{b_{p,p}^{(f)}[k] \cdot b_{p',p'}^{(f)}[k]}}, \tag{14}$$

[1] This is a model choice, and does not specifically account for the type of signal feature.

where the $b_{p,p'}^{(f)}[k]$ are obtained from the mean vector $\boldsymbol{\mu}^{(f)}[k]$ by reversing the vectorization (see (3) and (4))

$$\mathbf{B}^{(f)}[k] = \mathrm{vech}^{-1}\left(\boldsymbol{\mu}^{(f)}[k]\right). \tag{15}$$

B. Channel Similarity Matrix

Similar to [11], the dominant left singular vector $\mathbf{g}_p[k]$ of

$$\mathbf{C}_p[k] = \begin{bmatrix} \tilde{b}_{p,1}^{(1)}[k] & \dots & \tilde{b}_{p,1}^{(F)}[k] \\ \vdots & & \vdots \\ \tilde{b}_{p,P}^{(1)}[k] & \dots & \tilde{b}_{p,P}^{(F)}[k] \end{bmatrix} \in \mathbb{R}^{P\times F} \tag{16}$$

defines the similarity vector of microphone p relative to all P microphones while accounting for all F features. Instead of computing a costly SVD as in [11] in each time step, we recursively update $\mathbf{g}_p[k]$ over time to reduce computational load. Note that the left singular vectors of $\mathbf{C}_p[k]$ are identical to the eigenvectors of $\mathbf{C}_p[k]\mathbf{C}_p^{\mathrm{T}}[k]$, such that the desired dominant left singular vector is readily obtained from the *power method* [15]. While more iterations produce more accurate estimates, our experiments indicate that the spectrum of $\mathbf{C}_p[k]$ varies slowly, such that a single iteration per time step is sufficient to track $\mathbf{g}_p[k]$. Hence, the recursive update is [15]

$$\tilde{\mathbf{g}}_p[k+1] = \left(\mathbf{C}_p[k]\mathbf{C}_p[k]^{\mathrm{T}}\right)\mathbf{g}_p[k], \tag{17}$$

$$\mathbf{g}_p[k+1] = \frac{\tilde{\mathbf{g}}_p[k+1]}{\|\tilde{\mathbf{g}}_p[k+1]\|_2}. \tag{18}$$

Once all $\mathbf{g}_p[k],\ \forall p \in \{1,\dots,P\}$ are updated, the overall channel similarity matrix $\mathbf{R}[k]$ is constructed as in [11] by concatenation of the normalized dominant singular vectors

$$\mathbf{R}[k] = \begin{bmatrix} \frac{\mathbf{g}_1[k]}{[\mathbf{g}_1[k]]_1} & \dots & \frac{\mathbf{g}_P[k]}{[\mathbf{g}_P[k]]_P} \end{bmatrix} \in \mathbb{R}^{P\times P}. \tag{19}$$

C. Fiedler Vector

In [11], the coherence between pairs of microphone signals is modeled as an undirected, weighted graph $\mathcal{G}(\mathcal{V},\mathcal{E})$. The vertices $\mathcal{V}$ represent the microphones while the weights of the edges $\mathcal{E}$ reflect the pair-wise similarity of the microphone signals. The microphone utility is then obtained as the *Fiedler* vector [16], denoted by $\mathbf{u}[k]$, of $\mathcal{G}$, i. e., the eigenvector associated with the smallest non-zero eigenvalue of the random-walk graph Laplacian $\mathbf{L}$ [17]. As $\mathcal{G}$ is undirected, its weighted connectivity matrix $\mathbf{W}[k]$ is symmetric [18]. However, since $\mathbf{R}[k]$ in (19) is generally not symmetric, we choose

$$\left[\mathbf{W}[k]\right]_{pp'} = \left[\mathbf{W}[k]\right]_{p'p} = \frac{1}{2}\left(\left|\left[\mathbf{R}[k]\right]_{pp'}\right| + \left|\left[\mathbf{R}[k]\right]_{p'p}\right|\right). \tag{20}$$

As strongly negatively correlated features hint at a functional relation just like strongly positively correlated features, taking the absolute value in (20) ensures that channels with negatively correlated features are also considered similar.

As an eigenvector, the Fiedler vector $\mathbf{u}[k]$ is only unique up to scaling, especially by -1, i. e., negation. Thus, without additional side information, it is unclear whether useful microphones exhibit high or low values in $\mathbf{u}[k]$, such that considering only the magnitude of the elements in the Fiedler vector $\mathbf{u}[k]$ is meaningless. In [11], this sign ambiguity was resolved by relating the elements of $\mathbf{u}[k]$ to the degree matrix, following the assumption that well-connected nodes indicate useful microphones since the SoI evokes similar features in each channel. However, even statistically independent additive noise signal components can produce similar signal features if their statistics are similar. If the SoI is strongly attenuated in the majority of channels, e. g., due to large source-microphones distances, the cause of similar features may no longer be the SoI but rather the noise statistics. To avoid this problem, here, we resolve this ambiguity by preferring structured over unstructured signals as captured by their differential entropy [19] instead. This choice is justified here by considering that typical SoI signals like speech are highly structured and their entropy is low, while typical degradations like reverberation or additive noise are less structured and thus have higher entropy. Therefore, in each time step, we negate the Fiedler vector $\mathbf{u}[k]$ whenever the Pearson correlation coefficient of the estimated $\mathbf{u}[k]$ and the vector of differential entropies of all microphone signals

$$\mathbf{h}[k] = \begin{bmatrix} h_1[k] & \cdots & h_P[k] \end{bmatrix}^{\mathrm{T}} \tag{21}$$

is positive, such that high utility values correspond to low entropy and vice versa.

III. Experimental Evaluation

In this section, we illustrate the efficacy of the proposed online microphone utility estimation scheme using synthesized and recorded data. We provide the common algorithmic parameters for all experiments and define the performance measures used for evaluation, before moving on to the actual experiments.

A. Parameters and Performance Measures

The microphone signals are sampled at $f_{\mathrm{s}} = 16\,\mathrm{kHz}$ and characterized using $F = 3$ features: the skewness and kurtosis of the time-domain waveform and spectral flux [12]. The features are extracted from signal blocks of length $L = 1024$ samples with a shift of $M = 512$ samples between successive blocks, i. e., $50\,\%$ overlap. Signal entropy is estimated using a histogram-based approach from longer signal blocks ($32\,000$ samples $\approx 2\,\mathrm{s}$ with a shift of 512 samples) for more robust estimates, but is only used for resolving the sign ambiguity and not to characterize the signals. The temporal averaging factor for $\overline{\mathbf{a}}^{(f)}[k]$ is $\lambda = 0.99$ and the noise scaling factors for the KF in (8) and (9) are chosen heuristically based on experimental results as $\alpha_1 = 10^{-4}$ and $\alpha_2 = 0.2$, respectively.

As ground truth for the utility of the p-th microphone signal, we employ the frequency-averaged Magnitude-Squared Coherence (MSC) between the source signal and the p-th microphone signal, i. e.,

$$\gamma_p[k] = \frac{1}{N}\sum_{\nu=1}^{N}\left|\frac{\hat{\Phi}_{\mathrm{src},p}[k,\nu]}{\sqrt{\hat{\Phi}_{\mathrm{src}}[k,\nu]\cdot\hat{\Phi}_p[k,\nu]}}\right|^2, \tag{22}$$

to remain application-agnostic. Here, $N = L$ is the Discrete Fourier Transform (DFT) length and $\hat{\Phi}_{\mathrm{src},p}[k,\nu]$, $\hat{\Phi}_{\mathrm{src}}[k,\nu]$ and $\hat{\Phi}_p[k,\nu]$ are short-time estimates of the cross-Power Spectral Density (PSD) and auto-PSDs of source and microphone signals, respectively. Collecting the ground-truth utilities for all channels in the vector

$$\boldsymbol{\gamma}[k] = \begin{bmatrix} \gamma_1[k] & \dots & \gamma_P[k] \end{bmatrix}^{\mathrm{T}}, \tag{23}$$

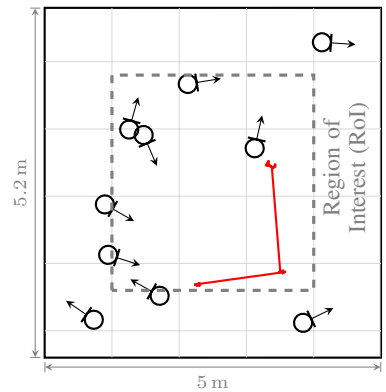

Fig. 1: Illustration of the experimental setup for synthesized data. Arrows indicate direction of maximum microphone sensitivity (cardioid pattern). A single, exemplary source trajectory is shown in red for illustration.

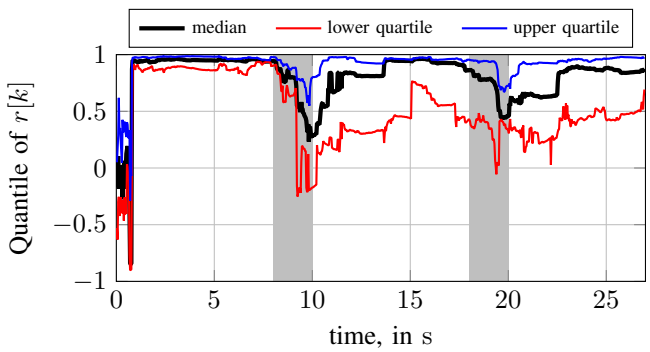

Fig. 2: Quantiles of Pearson correlation coefficient $r[k]$ between Fiedler and MSC vectors for synthesized data. Shaded areas indicate time intervals of source movement.

the performance of the proposed algorithm is measured by the Pearson correlation coefficient of $\mathbf{u}[k]$ and $\boldsymbol{\gamma}[k]$, i. e.,

$$r[k] = \frac{\left(\mathbf{u}[k] - \overline{u}[k]\mathbf{1}_P\right)^{\mathrm{T}}\left(\boldsymbol{\gamma}[k] - \overline{\gamma}[k]\mathbf{1}_P\right)}{\left\|\mathbf{u}[k] - \overline{u}[k]\mathbf{1}_P\right\|_2 \cdot \left\|\boldsymbol{\gamma}[k] - \overline{\gamma}[k]\mathbf{1}_P\right\|_2}, \tag{24}$$

where $\overline{u}[k]$, $\overline{\gamma}[k]$ denote the average of the elements of $\mathbf{u}[k]$, $\boldsymbol{\gamma}[k]$, respectively, and $\mathbf{1}_P$ is the all-ones vector of length P.

B. Synthesized data

The setup for synthesized data is depicted in Figure 1. A total of $P = 10$ cardioid microphones with the direction of maximum sensitivity indicated by arrows are placed in a room of dimensions $5\,\mathrm{m} \times 5.2\,\mathrm{m} \times 3\,\mathrm{m}$ with a reverberation time of $T_{60} = 500\,\mathrm{ms}$. A single SoI moving in the Region of Interest (RoI) and emitting a speech signal of $27\,\mathrm{s}$ duration is simulated using [20]. The source moves linearly to a new random position from $8\,\mathrm{s}$–$10\,\mathrm{s}$, and from $18\,\mathrm{s}$–$20\,\mathrm{s}$, indicated by the shaded gray areas in Figure 2. During other time intervals, the source randomly moves in a $0.2\,\mathrm{m} \times 0.2\,\mathrm{m}$ square around a fixed position to simulate slight head movements of a human speaker without displacement of the body. All microphones and the source are located in a horizontal plane at a height of $1\,\mathrm{m}$. Additive white Gaussian noise of identical power is added to each microphone channel such that a long-term Signal-to-Noise Ratio (SNR) of $10\,\mathrm{dB}$ is obtained at the microphone with the strongest SoI source image. Due to the source movement, the SNR values can vary drastically if the source comes very close to one of the microphones with typical values in the range of $5\,\mathrm{dB}$ to $10\,\mathrm{dB}$. The experiment is repeated for ten different realizations of the random source trajectory and two different speech signals (one male and one female) for a total of $R = 20$ trials.

Figure 2 shows the time-varying median of $r[k]$ across all trials in black, as well as the lower and upper quartile in red and blue, respectively. Ideally, $r[k]$ should quickly approach a value of 1 and remain there for the entire duration of the experiment. Evidently, the proposed algorithm initially produces utility estimates that agree very well with the observed coherence after only $1\,\mathrm{s}$. Although the agreement rapidly deteriorates during strong source movement, indicated by the shaded gray time intervals in Figure 2, the median shows that $r[k]$ re-attains values of 0.8 within about $3\,\mathrm{s}$ in the majority of trials. Across all signal blocks and trials, $r[k]$ achieves a mean value of $m_r = 0.666$ with a variance of $\sigma_r^2 = 0.426$.

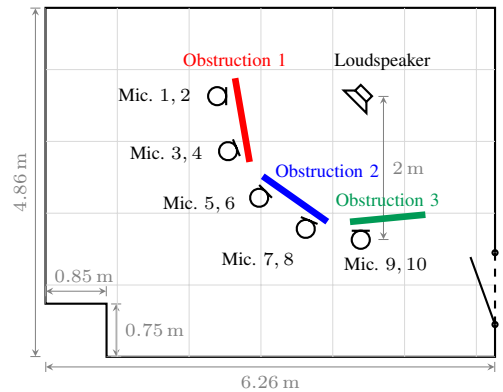

Fig. 3: Illustration of the experimental setup for recorded data. Line-of-sight between the loudspeaker and microphones 1–4 is blocked.

C. Recorded data

Figure 3 illustrates the experimental setup using recorded data. As for synthesized data, $P = 10$ microphones are placed in a recording room of dimensions $6.26\,\mathrm{m} \times 4.86\,\mathrm{m} \times 3\,\mathrm{m}$ with a reverberation time of $T_{60} \approx 320\,\mathrm{ms}$. The microphones are arranged in five pairs with inter-microphone distance $4\,\mathrm{cm}$ placed on a quarter circle of $2\,\mathrm{m}$ radius around a loudspeaker representing the SoI. The loudspeaker and the microphones are located in a horizontal plane at a height of $1.17\,\mathrm{m}$. To replicate a realistic scenario with physical obstacles, solid wooden panels $1\,\mathrm{m}$ wide and $2\,\mathrm{m}$ high are placed between the microphones and the loudspeaker, thereby suppressing the direct-path contribution in the corresponding microphone signal. The SoI signal consists of $51\,\mathrm{s}$ of speech, which is split uniformly into three segments of $17\,\mathrm{s}$ duration. During each segment, the corresponding obstacle is present while the other

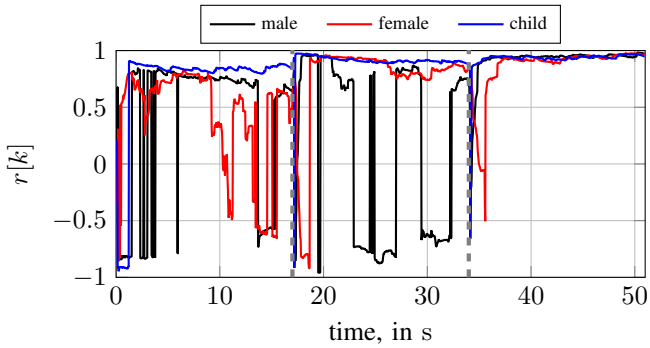

Fig. 4: Pearson correlation coefficient $r[k]$ between Fiedler and MSC vectors for recorded data. Dashed lines indicate changes of the obstruction's position.

obstacles are absent, i. e., in segment 1 (0 s–17 s) only obstacle 1 is present, etc. Three trials, each with a different signal (adult male speech, adult female speech, children's speech), are conducted. As for the synthesized data, white Gaussian noise is added to the microphone channels to obtain a long-term SNR of 10 dB at the microphone with the strongest SoI source image. Since all microphones are facing the SoI, the SNR varies less between microphones relative to the scenario with synthesized data above. As a result, the evoked signal features for the Fiedler vector computation as well as the signal entropy for resolving the sign ambiguity are less discriminative. Therefore, a performance degradation relative to the previous experiment must be expected.

Figure 4 shows $r[k]$ over time for the three different excitation signals. Performance is best for children's speech consistently achieving values of $r[k] > 0.8$ and re-converging almost instantly when the position of the obstacle changes. For female speech, re-convergence after the obstruction changes position takes around 2 s–3 s. Although the performance during the second and third segment is high, there are problems in the second half of the first segment. For male speech, although the Fiedler vector is estimated accurately, the proposed algorithm struggles with resolving the sign ambiguity of the Fiedler vector as indicated by the rapid sign changes of $r[k]$ in the first two segments.

IV. Conclusion

In this contribution, we expanded our prior work in [10], [11] to facilitate online estimation of the microphone utility in time-variant scenarios, which constitutes a highly relevant practical problem in WASNs. For an efficient online implementation, we introduced a KF formulation to recursively estimate feature correlation coefficients and introduced recursive updates for the similarity vector of each microphone channel. Furthermore, we proposed an improved approach to robustly resolve the sign ambiguity of the Fiedler vector based on the entropy of the recorded microphone signals. Experiments using synthesized and recorded data demonstrated the efficacy of the proposed method for tracking time-varying microphone utility in realistic acoustic conditions.

Future work includes verifying the practical efficacy of the proposed microphone ranking approach for different classes of online signal processing algorithms such as multichannel signal enhancement and localization of acoustic sources. Further improvements to even more robustly resolve the sign ambiguity of the Fiedler vector in realistic scenarios presents another research direction of considerable practical interest.

References

[1] A. Bertrand and M. Moonen, "Distributed adaptive estimation of correlated node-specific signals in a fully connected sensor network," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 2053–2056.

[2] A. Brendel and W. Kellermann, "Distributed source localization in acoustic sensor networks using the coherent-to-diffuse power ratio," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 1, pp. 61–75, 2019.

[3] A. Bertrand, "Applications and trends in wireless acoustic sensor networks: A signal processing perspective," in *2011 18th IEEE Symposium on Communications and Vehicular Technology in the Benelux (SCVT)*, Nov. 2011, pp. 1–6.

[4] L. Griffiths and C. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Transactions on Antennas and Propagation*, vol. 30, no. 1, pp. 27–34, Jan. 1982.

[5] S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov, "A Consolidated Perspective on Multimicrophone Speech Enhancement and Source Separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 4, pp. 692–730, Apr. 2017.

[6] K. Kumatani, J. McDonough, J. Lehman, and B. Raj, "Channel selection based on multichannel cross-correlation coefficients for distant speech recognition," in *Joint Workshop Hands-free Speech Commun. Microphone Arrays (HSCMA)*, Edinburgh, UK, May 2011, pp. 1–6.

[7] A. Bertrand and M. Moonen, "Efficient sensor subset selection and link failure response for linear MMSE signal estimation in wireless sensor networks," in *European Signal Process. Conf. (EUSIPCO)*, Aalborg, Denmark, Aug. 2010, pp. 1092–1096.

[8] J. Szurley, A. Bertrand, P. Ruckebusch, I. Moerman, and M. Moonen, "Greedy distributed node selection for node-specific signal estimation in wireless sensor networks," *Signal Processing*, vol. 94, pp. 57–73, Jan. 2014.

[9] M. Wolf and C. Nadeu, "Towards microphone selection based on room impulse response energy-related measures," in *Proc. of I Joint SIG-IL/Microsoft Workshop Speech Lang. Technol. Iberian Lang.*, Porto Salvo, Portugal, 2009, p. 4.

[10] M. Günther, A. Brendel, and W. Kellermann, "Single-channel signal features for estimating microphone utility for coherent signal processing," in *Proc. 23rd International Congress on Acoustics*, Sep. 2019, pp. 2716–2723.

[11] M. Günther, H. Afifi, A. Brendel, H. Karl, and W. Kellermann, "Network-aware optimal microphone channel selection in wireless acoustic sensor networks," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Jun. 2021.

[12] G. Peeters, "A large set of audio features for sound description (similarity and classification)," *CUIDADO project Ircam technical report*, 2004. [Online]. Available: http://recherche.ircam.fr/equipes/analyse-synthese/peeters/ARTICLES/Peeters_2003_cuidadoaudiofeatures.pdf

[13] T. Virtanen, M. Plumbley, and D. Ellis, *Computational analysis of sound scenes and events*. Springer, 2018.

[14] H. V. Henderson and S. R. Searle, "Vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics," *Canadian Journal of Statistics*, vol. 7, no. 1, pp. 65–81, 1979.

[15] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. The Johns Hopkins University Press, 2013.

[16] M. Fiedler, "Algebraic connectivity of graphs," *Czechoslovak Mathematical Journal*, vol. 23, no. 2, pp. 298–305, 1973.

[17] F. R. Chung and F. C. Graham, *Spectral graph theory*. American Mathematical Soc., 1997, no. 92.

[18] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.

[19] T. M. Cover and J. A. Thomas, *Elements of information theory*, 2nd ed. Hoboken, N.J: Wiley-Interscience, 2006.

[20] E. A. P. Habets, "Signal generator for MATLAB," Sep. 2011. [Online]. Available: https://github.com/ehabets/Signal-Generator