A Coherence-based Clustering Method for Multichannel Speech Enhancement in Wireless Acoustic Sensor Networks

Antonio J. Muñoz-Montoro
Computer Science Department
Universidad de Oviedo
Gijón, Spain
munozantonio@uniovi.es

Pedro Vera-Candeas
Telecommunication Engineering Department
Universidad de Jaén
Linares, Spain
pvera@ujaen.es

Mads Græsbøll Christensen
Audio Analysis Lab, CREATE
Aalborg University
Aalborg, Denmark
mgc@create.aau.dk

***Abstract*—Speech enhancement constitutes a great challenge in unknown noisy environments. Many studies have addressed this problem for both single-channel and centralized multichannel cases. However, in a real-world scenario, the effect of the reverberation and interference sounds degrades the performance of the state-of-the-art methods. In this sense, speech and signal processing with wireless acoustic sensor networks (WASNs) is becoming more and more popular, since they are able to physically cover a larger space and capture more spatial information mitigating the effect of the reverberation and the interference. In this paper, we present an unsupervised clustering method to cluster the nodes in a WASN into subnetworks, which detect different speakers. Thus, each subnetwork will be interested in detecting one of the multiple speakers in the acoustic scene. The proposed node clustering is based on the estimation of the magnitude-squared coherence between microphones observations, which measures the degree of their linear dependency. Then, a non-negative matrix factorization (NMF) based approach is developed and applied to find the optimal clustering. Simulation results show that the proposed clustering method can assign nodes into subnetworks based on the microphones observations obtaining promising results.**

***Index Terms*—WASN, node clustering, coherence, NMF, accuracy, confusion matrix**

I. Introduction

Multichannel signals are typically recorded by using microphone arrays. This enables to exploit spatial diversity and allows to localize target sound sources and/or to cancel out interfering sound sources coming from certain directions [1]–[3]. Several application using microphone arrays can be found in the literature, e.g., automatic speech recognition, hearing aids, computer games, teleconferencing systems, hands-free telephony, etc.

Conventional microphone arrays have limitations and provide a low performance in certain contexts [4], [5]. Such arrays only sample the sound field locally and often at a relatively large distance from the target sources. Moreover, in the case of portable devices, these systems are limited by space and power constraints.

As an alternative, wireless acoustic sensor networks (WASNs) provide many advantages. A WASN is comprised by several acoustic nodes, where each node can be formed by a single microphone or by a conventional microphone array. These nodes are randomly placed in the acoustic environment and are interconnected in an ad hoc fashion, providing a great scalability and versatility. This allows them to physically cover a much larger area, which increases the probability to have a subset of microphones close to target sound sources, obtaining higher quality recordings. In addition, note that the size limitations of the arrays are relaxed because of wireless interconnection. Due to these benefits, WASNs has become in an emerging topic that has attracted much attention from the signal processing community [5], [6].

For the design of audio signal processing applications in WASNs, two different network topologies can be exploited, i.e., centralized and distributed. The former assumes that the information of all nodes compounding the network is accesible at one central node. In this case, optimal solutions can be obtained by managing this information properly. On the other hand, the distributed methods process signals locally in each node, without the requirement of a fusion center. Thus, the large communication bandwidth requirements and the long distance communication are reduced as every node only needs to communicate and exchange information with its neighbors [4], [5]. Moreover, the distributed topology allows to distribute the computational burden over the WASN, reducing the amount of data processing as in a fusion center with a centralized method [4].

Several methods in the literature have addressed both centralized and distributed approaches for different tasks. Souden et al. [7] proposed a multichannel noise tracking method to estimate the multichannel speech presence probability. The experimentation revealed that the speech detection performance improves when the microphone number increases. Note that the proposed system only operates in a centralized manner. Although the obtained results were promising, the

Part of the research leading to these results has received funding from the Regional Government of Andalucía under the "Proyectos de I+D+i del Plan Andaluz de Investigación, Desarrollo e Innovación (PAIDI 2020)" Framework Programme through grant *P18-TPJ-4864*.

conclusion was that the noise tracking method requires a careful initialization. Following this work, Taseska and Habets [8] applied the multichannel noise tracking method to sound extraction by using distributed microphone arrays. Nevertheless, the proposal is still a centralized solution. Bahari et al. [9] used WASNs for the multi-speaker voice activity detection (VAD) problem. The authors proposed to form node clusters and compute the VAD for different speakers at each cluster. However, the proposed method presents a high computational burden, since the distributed eigenvalue decomposition (EVD) is required to obtain the node clusters. In [10], the authors developed a node clustering method based on the fuzzy c-Means algorithm. This method was subsequently used for source separation in ad hoc multichannel arrays [11]. A topology-independent distributed adaptive node-specific signal estimation algorithm was introduced in [12]. In this method, each node of the WASN is tasked with estimating a node-specific desired signal. The obtained results showed a robust behavior in the face of topology and scalability changes of the network. More recently, [5] developed a centralized and distributed model-based node clustering method based on the k-means algorithm to estimate the speech presence probability. The method clusters the nodes in the WASN so the entire network is divided into subnetworks. Each subnetwork is interested in detecting one of the multiple speakers in the acoustic scene.

WASNs have been also used for the particular case of sound field control applications, and more specifically for Active Noise Control (ANC) [13], [14]. Preliminary results showed that incorporating clustering strategies allows to improve the performance of algorithms developed in this field [15].

In this work, we propose an unsupervised clustering method for WASNs. We propose to use coherence between microphones observations to divide the WASN into subnetworks that detect different speakers. The magnitude-squared coherence measures the degree of the linear dependency of the microphones observations by analyzing similar frequency components. Subsequently, a non-negative matrix factorization (NMF) approach is applied, taking advantage of the clustering property inherent to this technique. In the end, the proposed method allows to perform the clustering dynamically with a very low computational burden, which makes it suitable for a multitude of audio signal processing applications.

The rest of this paper is organized as follows. Section II presents the signal model and the problem formulation. Section III discusses the proposed node clustering method. Section IV presents the experimentation and obtained results. Finally, conclusions are presented in Section V.

II. Problem formulation

The problem considered in this work is to cluster M microphones, randomly deployed in a room environment, into K clusters or subnetworks. Thus, each subnetwork will be focused on detecting one of the K speakers in the acoustic scene. Note that here each node in the WASN is a single microphone. Fig. 1 illustrates the problem for a simple network.

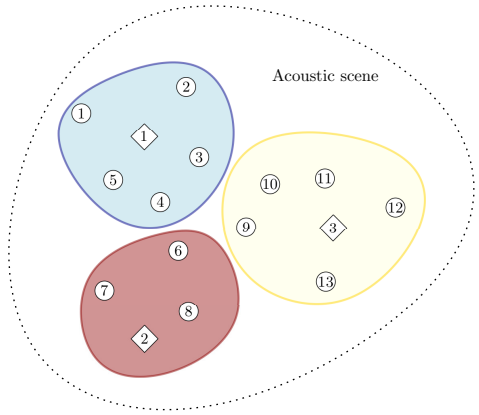


Fig. 1: A WASN of 13 nodes divided in 3 clusters. Each white circle represents a node and each diamond represents a speaker.

The observed signal $x_m(t)$ at the m-th microphone and time instant t can be expressed as

$$x_m(t) = s_m(t) + v_m(t), \tag{1}$$

where $s_m(t)$ is the clean speech and $v_m(t)$ is the noise signal plus interference. Collecting a frame of observation signal samples into a vector form, the linear signal model in (1) can be reformulated as

$$\begin{aligned}\mathbf{x}_m(t) &= [x_m(t)x_m(t-1)\dots x_m(t-T+1)]^{\mathrm{T}} \\ &= \mathbf{s}_m(t) + \mathbf{v}_m(t),\end{aligned} \tag{2}$$

where T denotes the frame size and the superscript ^{T} refers to the matrix transpose. Note that $\mathbf{s}_m(t)$ and $\mathbf{v}_m(t)$ are defined similarly to $\mathbf{x}_m(t)$ and denote the clean speech and noise signal vectors, respectively.

III. Proposed clustering algorithm

The proposed clustering method consists in two steps. The first step is to compute the magnitude-squared coherence between microphones observations in order to measure the degree of their linear dependency. Then, NMF is applied over the coherence process output to find the optimal clustering.

A. Spectral magnitude-squared coherence measure

The magnitude-squared coherence is a statistic that can be used to analyze the linear relationship between two audio signals $x(t)$ and $y(t)$ [16]. This statistic can be obtained by computing the fast Fourier transform (FFT) of both signals, and then by measuring coherence as a function of the center frequency of the filter. Therefore, the magnitude-squared coherence can be obtained as a frequency dependent function using:

$$\Gamma_{xy}(f) = \frac{|S_{xy}(f)|^2}{S_{xx}(f)S_{yy}(f)}, \tag{3}$$

where $S_{xy}(f)$ is the cross spectral density (CSD), which is really a spectral correlation density. The CSD can be computed as

$$S_{xy}(f) = \sum_{k=1-T}^{T-1} R_{xy}(k) e^{-i2\pi fk}, \tag{4}$$

where $R_{xy}(k)$ is the cross correlation function between $x(t)$ and $y(t)$ and T denotes the frame size. Note that here multiple short-time Fourier transform are averaged. Likewise, this cross correlation can be estimated by

$$R_{xy}(k) = \begin{cases} \frac{1}{T}\sum_0^{T-1-k} x(t)y(t+k) & k = 0, \ldots, T-1 \\ R_{xy}(-k) & k = -(T-1), \ldots, -1. \end{cases} \tag{5}$$

The coherence measure is a statistical indicator which indicates how two signals are correlated. Note that $\Gamma_{xy}(f) \in [0, 1]$, where the value 1 indicates a perfectly linear relationship between both signals, and the value 0 indicates a complete lack of correlation. The magnitude-squared coherence in (3) can be used to measure the correlation between all the signals captured by the WASN microphones. In order to consider the same weight to all frequency bins regardless of their power, we propose to compute the following coherence metric,

$$C_{xy} = \frac{\sum_{f=0}^{F} \Gamma_{xy}(f)}{F} \in [0, 1]. \tag{6}$$

Finally, arranging all the coherence measures between the audio observations, a non-negative symmetric coherence matrix $\mathbf{C} \in \mathbb{R}_+^{M\times M}$ can be obtained as,

$$\mathbf{C} = \begin{bmatrix} 1 & \cdots & \cdots & C_{1M} \\ C_{12} & 1 & \cdots & C_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1M} & C_{2M} & \cdots & 1 \end{bmatrix}, \tag{7}$$

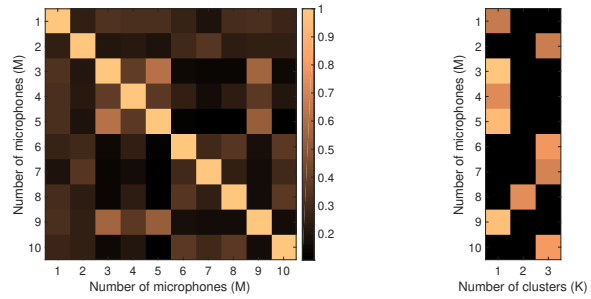
B. NMF-based model for clustering coherence observations

Here, we will briefly review the proposed NMF analysis applied over the coherence matrix $\mathbf{C}$ to obtain the optimal clusters.

Firstly, let us analyze the information contained in $\mathbf{C}$. The j-th row (or column) of $\mathbf{C}$ represents the degree of correlation between the audio observation captured by the j-th microphone and the rest of M observations. This way, the microphones which are closest to a specific source are highly correlated. Mathematically, $\mathbf{C}$ can be considered as a linear subspace of dimension M. Thus, the clustering process consists in downgrading this subspace into a linear subspace of dimension K. This may be achieved by taking into account the inherent clustering property of the NMF [17]. In addition, note that the NMF is well-suited because of the non-negativity of the coherence matrix.

For the NMF analysis, the non-negative symmetric coherence matrix $\mathbf{C}$ can be modeled as

$$\mathbf{C} = \mathbf{B}\mathbf{B}^{\mathrm{T}} \odot (\mathbf{1} - \mathbf{I}) + \mathbf{I} \tag{8}$$



(a) Coherence matrix ($\mathbf{C}$) (b) Cluster matrix ($\mathbf{B}$)

Fig. 2: Example of microphone clustering based on NMF. The number of clusters is 3 and the number of microphones is 10. The red rectangles indicate the microphones assigned to the first cluster.

where the $\odot$ denotes the Hadamard product. $\mathbf{B} \in \mathbb{R}_+^{M\times K}$ is the cluster matrix, $\mathbf{I} \in \mathbb{N}^{M\times M}$ is an identity matrix and $\mathbf{1} \in \mathbb{N}^{M\times M}$ is an all-ones matrix. Let us explain the model in (8). Due to the symmetric property of $\mathbf{C}$, the proper way to model it is by $\mathbf{B}\mathbf{B}^{\mathrm{T}}$. However, the main diagonal of $\mathbf{C}$ does not provide any relevant information in the learning process of $\mathbf{B}$. This is why we introduce $\mathbf{I}$ and $(\mathbf{1} - \mathbf{I})$.

Based on Euclidean divergence, $\mathbf{B}$ can be estimated using iterative multiplicative update rules [18]:

$$\mathbf{B} \leftarrow \mathbf{B} \odot \frac{(\mathbf{C} \odot (\mathbf{1} - \mathbf{I}))\mathbf{B}}{(\mathbf{B}\mathbf{B}^{\mathrm{T}} \odot (\mathbf{1} - \mathbf{I}))\mathbf{B}} \tag{9}$$

Note that the division is element-wise division operation.

For the clustering application, the columns of the cluster matrix $\mathbf{B}$ contain the contribution of each microphone to each cluster. The clustering result is then obtained by

$$\gamma_m = \{j \in [1, K] : b_{mj} \geq b_{mk}, \forall k \in [1, K]\}, \tag{10}$$

where γ_m denotes the cluster assigned to the m-th microphone and b_{mk} is the (m, k) entry of the matrix $\mathbf{B}$.

Fig. 2 shows an example of microphone clustering using the proposed NMF strategy. In this example, the objective is to cluster ten microphones into three subnetworks that are focused on three speakers. As can be observed in Fig. 2a, the column-vectors of the 3rd, 4th, 5th and 9th microphones keep a high correlation. Fig. 2b shows how this correlation has been identified by the NMF and how these microphones have been assigned to the 1st cluster. Therefore, we can assume that these microphones are close to the same speaker.

The node clustering method is summarized in Algorithm 1.

IV. Evaluation and Results

In this section, simulations are performed to show the performance of the proposed clustering method in simulated room acoustic environments.

A. Experimental Setup

The experimental evaluation was carried out using a database compounded by speech signals taken from the

Algorithm 1 Proposed clustering method

for $i = 1$ to M **do**
  for $j = i$ to M **do**
    Compute the cross correlation $R_{x_i x_j}(k)$ using (5).
    Compute the CSD $S_{x_i x_j}(f)$ using (4).
    Obtain the coherence measure $\Gamma_{x_i x_j}(f)$ using (3).
    Compute the coherence metric $C_{x_i x_j}$ using (6).
  end for
end for
Initialize $\mathbf{B}$ with random values.
for $iters = 1$ to $MaxIter$ **do**
  Update $\mathbf{B}$ according to (9).
end for
Obtain the optimal clustering by (10).

CHiME corpus [19] and noise signals extracted from the AURORA database [20]. The noise signals include babble, restaurant, exhibition noise, street and station noise. This background noise was applied equally to all microphones with SNR $= 10$ dB. Note that all the signals are downsampled to 8 kHz. The time-frequency representation is obtained using 2048-point STFT, and the frame size and the hop size for the STFT are set to 512 (64 ms) and 64 (8 ms) samples, respectively. We have simulated different mixing conditions using the image source model method [21] for a rectangular room of dimensions 10 m $\times$ 10 m $\times$ 3 m. Room impulse responses (RIR) were generated for reverberation times (T_{60}) of 200 ms and 400 ms, which provides moderate reverberation environments. We have simulated 50 nodes (microphones) randomly placed in the room, where the maximum distance was set to 2.5 m between them. Three speakers were located at $(8, 8, 1.5)$ m, $(6, 2, 1.5)$ m and $(3, 6, 1.5)$ m, respectively. The speech signals are scaled to have the same power before convolving with the RIRs.

For a quantitative evaluation of the reliability of our clustering method, we have used the accuracy metrics $Acc(\%)$, which can be defined as the percentage of microphones correctly assigned to their respective clusters. In this regard, the ground truth (GT) is computed from the RIR as

$$\gamma_m^{\mathrm{GT}} = \{j \in [1, K] : \sigma_{mj} \geq \sigma_{mk}, \forall k \in [1, K]\}, \qquad (11)$$

where $\sigma_{mk} = \sqrt{\sum_t h_{mk}^2(t)}$, being $h_{mk}(t)$ the spatial room impulse response of the speaker k captured by the microphone m. The GT gives us a measure to evaluate the reliability of the clustering procedure over the evaluated dataset.

We have also compared our proposal with the k-means-based clustering method presented in [5]. This method computes the clustering by initializing the algorithm with K cluster centers first. The authors use a feature $\mathbf{b}_m$ obtained by computing the Itakura-Saito (IS) divergence between the observation periodograms and a pretrained codebook of speech and noise AR models. The clustering result is then obtained by iterating between the following two steps: (1) feature $\mathbf{b}_m$ is assigned to its nearest cluster center $\mathbf{c}_k$ and (2) the cluster center $\mathbf{c}_k$ is then recomputed as the means of the data which is assigned to the k-th cluster. Iterating between step (1) and step (2) until convergence gives the final clustering result.

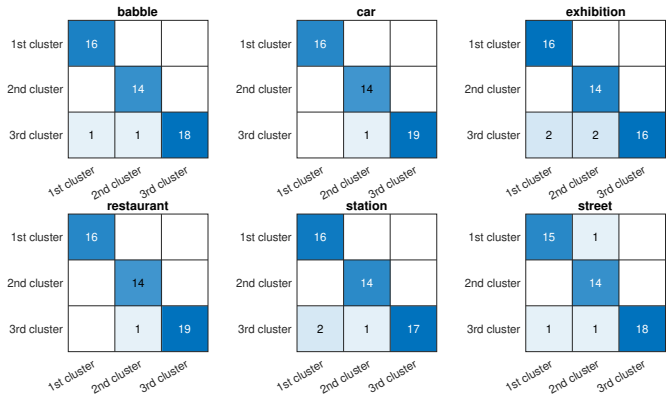


Fig. 3: Confusion matrices for the proposed method, obtained from the evaluation dataset for a reverberation time of $T_{60} =$ 200 ms. The number of clusters is 3 and the total number of microphones is 50.

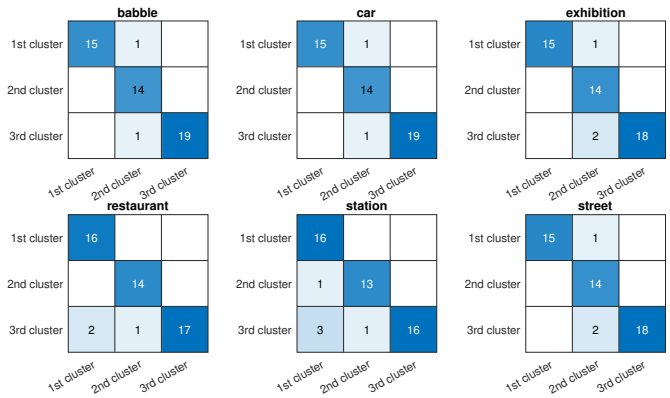


Fig. 4: Confusion matrices for the proposed method, obtained from the evaluation dataset for a reverberation time of $T_{60} =$ 400 ms. The number of clusters is 3 and the total number of microphones is 50.

B. Results

Fig. 3 and Fig. 4 show the confusion matrices for mixtures generated by using different types of background noise. For the problem described in this paper, the optimal number of clusters corresponds to the number of sources in the acoustic environment. In order to determine this number, we propose to use the variance ratio criterion (VRC) strategy as in [5]. In this case, the number of speakers is three and, therefore, three subnetworks are formed. Fig. 3 displays the clustering results obtained for $T_{60} = 200$ ms. In general, the performance across the different background noises types is very similar. It can be observed that the obtained clustering is very close to GT and only four microphones from the third cluster are assigned to the other two clusters in the worst case.

Fig. 4 shows the clustering results obtained for $T_{60} = 400$ ms. As can be observed, the method is robust against reverberation time and most of the microphones

are properly labelled to their respective clusters. The worst performance is obtained when the background noise is station. In this case, the method tends to include some microphones from the third clusters in the first cluster. Although to a lesser extent, a similar behavior can be observed for the restaurant, exhibition and street noise.

Table I reports the accuracy results provided by both methods when evaluating the database described in Section IV-A. Overall, the proposed method outperforms the method based on k-means and AR models. Note that the proposed method improves accuracy when the number of microphones is increased. This is because more spatial information is captured when using more microphones and NMF is able to generalize the clustering problem properly. Using more spatial information from the acoustic environment makes it easier to discriminate coherence patterns. This behaviour is not observed in the compared method. A worse performance of the proposal can be seen for restaurant noise compared to the k-means method when the reverberation time is 400 ms. This may be due to the use of pre-learned AR models.

TABLE I: Comparison of the clustering methods in terms of accuracy for each type of background noise and as a function of the microphone number and reverberation time.

Nº nodes	$T_{60}(ms)$	Method	Babble	Exhibition	Car	Restaurant	Station	Street
20	200	NMF	**95%**	**90%**	**100%**	85%	**90%**	**90%**
		K-means	85%	70%	85%	85%	80%	85%
	400	NMF	**85%**	**95%**	**85%**	85%	**90%**	**95%**
		K-means	80%	75%	80%	**95%**	80%	75%
50	200	NMF	**96%**	**92%**	**98%**	**98%**	**94%**	**94%**
		K-means	84%	74%	80%	78%	78%	74%
	400	NMF	**96%**	**94%**	**96%**	**94%**	**90%**	**94%**
		K-means	82%	74%	80%	70%	74%	77%

V. Conclusions

In this paper, we presented an unsupervised clustering method to cluster the nodes in a WASN into subnetworks, which detect different speakers. In particular, the proposed node clustering was based on the magnitude-squared coherence estimation between microphones observations. Then, a NMF strategy has been developed to find the optimal clustering. At the end, each subnetwork is focused on detecting one of the multiple speakers in the acoustic scene. The developed proposal allows to perform the clustering without any prior information of the speakers of the acoustic scenes. The results showed improved clustering performance in comparison to state-of-the-art method. Specifically, we reached an increase of 10% in terms of accuracy.

As future work, we would investigate a way to extend the proposal to a distributed network. Additionally, we will investigate the detection of the number of speakers within the proposed model.

References

[1] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*. 2001.

[2] S. Markovich, S. Gannot, and I. Cohen, "Multichannel Eigenspace Beamforming in a Reverberant Noisy Environment With Multiple Interfering Speech Signals," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, pp. 1071–1086, 8 2009.

[3] J. R. Jensen, M. G. Christensen, and A. Jakobsson, "Harmonic minimum mean squared error filters for multichannel speech enhancement," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 501–505, IEEE, 3 2017.

[4] 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)*, pp. 1–6, IEEE, 11 2011.

[5] Y. Zhao, J. K. Nielsen, J. Chen, and M. G. Christensen, "Model-based distributed node clustering and multi-speaker speech presence probability estimation in wireless acoustic sensor networks," *The Journal of the Acoustical Society of America*, vol. 147, pp. 4189–4201, 6 2020.

[6] G. Zhang and R. Heusdens, "Distributed Optimization Using the Primal-Dual Method of Multipliers," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 4, pp. 173–187, 3 2018.

[7] M. Souden, J. Chen, J. Benesty, and S. Affes, "An Integrated Solution for Online Multichannel Noise Tracking and Reduction," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, pp. 2159–2169, 9 2011.

[8] M. Taseska and E. A. P. Habets, "Informed Spatial Filtering for Sound Extraction Using Distributed Microphone Arrays," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, pp. 1195–1207, 7 2014.

[9] M. H. Bahari, L. K. Hamaidi, M. Muma, J. Plata-Chaves, M. Moonen, A. M. Zoubir, and A. Bertrand, "Distributed multi-speaker voice activity detection for wireless acoustic sensor networks," 2017.

[10] S. Gergen, A. Nagathil, and R. Martin, "Classification of reverberant audio signals using clustered ad hoc distributed microphones," *Signal Processing*, vol. 107, pp. 21–32, 2 2015.

[11] S. Gergen, R. Martin, and N. Madhu, "Source Separation by Feature-Based Clustering of Microphones in Ad Hoc Arrays," in *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 530–534, IEEE, 9 2018.

[12] J. Szurley, A. Bertrand, and M. Moonen, "Topology-Independent Distributed Adaptive Node-Specific Signal Estimation in Wireless Sensor Networks," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, pp. 130–144, 3 2017.

[13] J. Plata-Chaves, A. Bertrand, and M. Moonen, "Incremental multiple error filtered-X LMS for node-specific active noise control over wireless acoustic sensor networks," in *2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, vol. 2016-Septe, pp. 1–5, IEEE, 7 2016.

[14] M. Ferrer, M. de Diego, G. Pinero, and A. Gonzalez, "Affine Projection Algorithm Over Acoustic Sensor Networks for Active Noise Control," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 448–461, 2021.

[15] J. Plata-Chaves, A. Bertrand, M. Moonen, S. Theodoridis, and A. M. Zoubir, "Heterogeneous and Multitask Wireless Sensor Networks—Algorithms, Applications, and Challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, pp. 450–465, 4 2017.

[16] W. A. Gardner, "A unifying view of coherence in signal processing," *Signal Processing*, vol. 29, pp. 113–140, 11 1992.

[17] C. Ding, X. He, and H. D. Simon, "On the Equivalence of Nonnegative Matrix Factorization and Spectral Clustering," in *Proceedings of the 2005 SIAM International Conference on Data Mining*, (Philadelphia, PA), pp. 606–610, Society for Industrial and Applied Mathematics, 4 2005.

[18] D. D. Lee and H. S. Seung, "Algorithms for Non-negative Matrix Factorization," pp. 556–562, 2001.

[19] H. Christensen, J. Barker, N. Ma, and P. Green, "The CHiME corpus: A resource and a challenge for computational hearing in multisource environments," in *Proceedings of the 11th Annual Conference of the International Speech Communication Association, INTERSPEECH 2010*, pp. 1918–1921, 2010.

[20] D. Pearce and H. G. Hirsch, "The AURORA experimental framework for the performance evaluation of speech recognition systems under noisy conditions," in *6th International Conference on Spoken Language Processing, ICSLP 2000*, 2000.

[21] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, pp. 943–950, 4 1979.