

 Open access • Proceedings Article • DOI:10.1109/SCVT.2011.6101302

Applications and trends in wireless acoustic sensor networks: A signal processing perspective — [Source link](#)

Alexander Bertrand

Institutions: Katholieke Universiteit Leuven

Published on: 15 Dec 2011 - Symposium on Communications and Vehicular Technology in the Benelux

Topics: Wireless sensor network, Wireless microphone, Microphone and Signal processing

Related papers:

- [Image method for efficiently simulating small-room acoustics](#)
- [Distributed Adaptive Node-Specific Signal Estimation in Fully Connected Sensor Networks—Part I: Sequential Node Updating](#)
- [Beamforming: a versatile approach to spatial filtering](#)
- [Distributed Node-Specific LCMV Beamforming in Wireless Sensor Networks](#)
- [Reduced-Bandwidth and Distributed MWF-Based Noise Reduction Algorithms for Binaural Hearing Aids](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/applications-and-trends-in-wireless-acoustic-sensor-networks-2xqp86ruje>

Applications and trends in wireless acoustic sensor networks: a signal processing perspective

(Invited Paper)

Alexander Bertrand

Katholieke Universiteit Leuven - Dept. ESAT
Kasteelpark Arenberg 10, B-3001 Leuven, Belgium
E-mail: alexander.bertrand@esat.kuleuven.be; marc.moonen@esat.kuleuven.be

Abstract—Wireless microphone networks or so-called wireless acoustic sensor networks (WASNs) are a next-generation technology for audio acquisition and processing. As opposed to traditional microphone arrays that sample a sound field only locally, often at large distances from the relevant sound sources, WASNs allow to use many more microphones to cover a large area of interest. However, the design of such WASNs is very challenging, especially for real-time audio acquisition and signal enhancement due to the significant data traffic in the network. There is a need for scalable solutions, both on the signal processing level and on the network-communication level. In this paper, we give an overview of applications and trends in the field of WASNs, and we address the core challenges that need to be tackled. We mainly focus on the signal processing level, and we explain how advances in the area of signal processing can relax the high-demanding constraints on the network layer design. Furthermore, we address the interaction between the application layer and the network layer, and we explain why cross-layer design can be important to improve the performance of WASN applications.

I. INTRODUCTION

Microphone arrays (see Fig. 1) become more and more popular for audio acquisition, since multi-microphone recordings enable to exploit spatial diversity, allowing to localize target sound sources and/or to cancel out interfering sound sources coming from certain directions [1]–[4]. Microphone arrays are used in several applications, e.g., hearing aids, teleconferencing systems, hands-free telephony, automatic speech recognition, computer games, etc. [1].

Despite the obvious advantages over single-microphone systems, traditional microphone arrays still have their limitations and are often not sufficiently performant. Since a microphone array only samples the sound field locally, often at a relatively large distance from the target source(s), the recorded signals often have a low signal-to-noise ratio (SNR). Furthermore, due to obvious space and power constraints, especially in portable devices, the array is limited in physical size and in processing power. For example, only two or three microphones fit in a hearing aid, and the available power is limited due to the small batteries, which also limits the number of audio channels that can be processed by the device. However, it is common knowledge that the performance of microphone arrays improves when using more microphones, preferably at large inter-microphone distances.

To overcome these limitations, wireless microphone nodes, containing a single microphone or a small microphone array, can be

The work of A. Bertrand was supported by a Postdoctoral Fellowship of the Research Foundation - Flanders (FWO). This 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 (DYSCO, ‘Dynamical systems, control and optimization’, 2007-2011), 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.

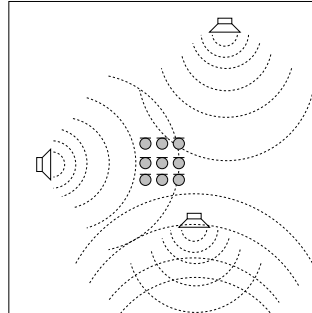


Fig. 1. Schematic example of a localized regularly arranged microphone array.

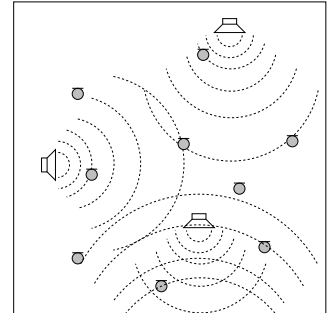


Fig. 2. Schematic example of a randomly distributed microphone array (WASN).

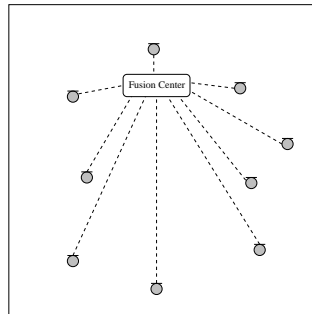


Fig. 3. Schematic example of centralized processing by means of a fusion center.

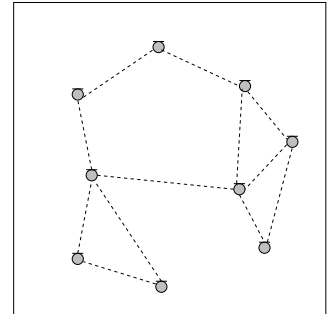


Fig. 4. Schematic example of distributed processing in a WASN with an ad hoc topology.

distributed randomly over the environment (see Fig. 2). This results in a wireless network of microphones or a so-called wireless acoustic sensor network (WASN). Due to the wireless communication, the array-size limitations disappear and the microphones can be placed at positions where it is difficult to place wired microphones. Furthermore, the microphone nodes physically cover a much larger area, which increases the probability to have a subset of microphones close to a sound source, yielding higher quality recordings. Because of these advantages, and since small microphones can now be produced at low cost, it is believed that WASNs will become very popular for audio acquisition and audio processing in the near future.

In some applications, the nodes of a WASN can transmit their recorded microphone signal(s) to a dedicated device (the fusion center or FC) where all signals are processed, resulting in a network with a centralized or star topology (see Fig. 3). However, in many applications such an FC is either unavailable, too far away from

certain nodes, or the total number of microphone signals is too large to process in a single device. In-network processing can then be a solution, i.e., the nodes can locally process data and share the result with their neighboring nodes, rather than with an FC (see Fig. 4). Such a distributed approach is often preferred, especially so when it is scalable in terms of communication bandwidth requirements and computational complexity. However, the algorithm design for such distributed settings is much more challenging, i.a., because each node only has access to a subset of the available data.

In general, all WASN applications, problem statements or algorithms can be classified into either *signal* estimation or *parameter* estimation techniques [5]. In the case of *signal* estimation (also referred to as signal enhancement), the goal is to estimate a desired signal (e.g., a speech signal), while suppressing background noise and/or removing reverberant components. This usually relies on fusion of the recorded signals at different nodes (see Fig. 5), requiring transmission of audio signals. In the case of *parameter* estimation, the goal is to extract certain parameters from the recorded audio signal(s), such as the location or identity of speakers, the acoustic properties of a room, or speech features. In this case, the nodes may only exchange parameter vectors or energy measurements at a slow time-scale compared to the sampling rate of the microphones. In this paper, we mainly focus on the former class (signal estimation), where the nodes actually transmit audio signals, rather than parameters¹. The real-time processing and streaming of audio data imposes challenging demands on the network layer with respect to data rate, synchronization, input-output (IO) delay and quality of service (QoS). These are typical requirements in the general class of so-called wireless multimedia sensor networks [6] (which also covers WASNs).

This paper gives an overview of the state of the art, the current trends, and future directions in digital signal processing (DSP) algorithms for WASNs. We focus both on the applications and on the enabling DSP techniques, without going into too much detail on the algorithms. We will identify the core challenges with respect to the algorithm design for such WASNs. Although we approach these challenges from a signal processing perspective (i.e., the application layer), most of them also apply to the network layer design. We explain how advances in the area of signal processing can relax the high-demanding specifications at the network layer. Furthermore, we explain why cross-layer design can be important to improve the performance of WASN applications.

II. EXAMPLE APPLICATIONS

In this section, we briefly address some example applications that could benefit from using WASNs.

- 1) **Hearing aids:** Reduction of acoustic background noise is crucial in hearing aids (HAs) to provide intelligible speech signals in noisy environments [7]. This noise reduction is usually obtained by using a local microphone arrays in the HA itself [1]. If a HA is worn at both ears, these can be connected with each other through a wireless link, yielding a so-called binaural HA [8], [9], which is essentially a 2-node WASN. By exchanging microphone signals between HAs, the noise reduction can be greatly improved since more (and well-separated) microphones can be used by each HA. Furthermore, systems exist where a remote wireless microphone is connected to a HA [7]. Since microphones become smaller and cheaper,

¹It is noted that some parameter estimation algorithms also require exchange of full-rate audio signals between nodes, e.g., for localization based on time difference of arrival.

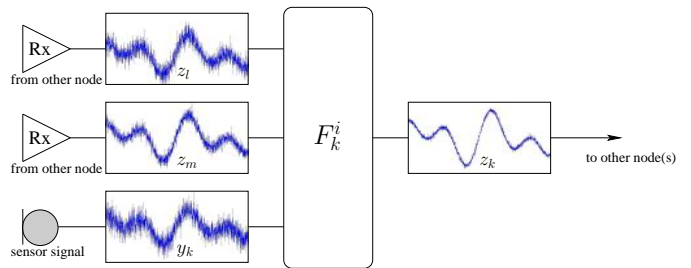


Fig. 5. Schematic illustration of signal estimation in node k , based on fusion of its microphone signal y_k with the signals z_l and z_m obtained from neighboring nodes l and m .

it is expected that this will evolve towards more complex WASNs with many more microphones, e.g., incorporated in clothing, furniture, or strategically placed by the user [5], [10]. Furthermore, since we are nowadays surrounded by wireless devices that are equipped with microphones (e.g. smartphones, laptops, hands-free kits, etc.), these can also be incorporated in the network to improve noise reduction in HAs.

- 2) **Hands-free telephony:** Noise reduction with microphone arrays is common in hands-free telephony, e.g., in cars or video conferencing. The use of additional microphone nodes (e.g. from tablets, smartphones, or dedicated devices) allows a significant improvement in the enhancement of the recorded speech signals [5], [11]–[14]. Furthermore, WASNs are an enabling technology for speech communication in noisy and dynamic environments such as airports, factories, stock markets, etc. The deployment of a WASN over the entire building allows people to walk freely through the building while talking on the phone, and benefit from the microphones in their neighborhood to enhance the recorded speech signal.
- 3) **Acoustic monitoring:** Currently, most literature on WASNs is in the context of acoustic monitoring of an environment, e.g., for vehicle tracking or classification, surveillance, etc. (see [2], [15], [16], and references therein). Compared to cameras, microphones are cheaper, and have the important advantage that they have no line-of-sight constraints. Video-based surveillance can also be integrated with a WASN [15].
- 4) **Ambient intelligence:** The term ‘ambient intelligence’ [17] refers to an intelligent environment that is aware of the presence of a user and that is responsive to its needs. The sensors and processors are wirelessly connected with each other and assumed to be inconspicuously incorporated in the environment. Active communication between the user and the environment will likely be based on automatic speech recognition (ASR), where signal enhancement is very important, since background noise and reverberation are known to significantly affect the recognition performance. A WASN with signal enhancement technology is able to acquire intelligible speech, wherever the speaker is positioned in the room. WASNs for ambient intelligence also use acoustic monitoring techniques for event detection, sound classification, localization, and speaker identification.

III. CORE CHALLENGES

In the design of signal processing algorithms for WASNs, we can define some core challenges that are absent in traditional (wired) microphone arrays. It is noted that energy awareness is not incorporated

in this list, since it is perceived here as a general goal to which many of the below challenges contribute.

- 1) **Unknown Array geometry:** In many cases, the positions of the microphone nodes in a WASN are not known a priori, due to the random deployment. For some tasks, such as localization or speech enhancement based on spatial separation (beamforming), supporting algorithms may be required to estimate node and/or source positions. In the context of WASNs, blind algorithms that do not require this extra information are usually preferred (see, e.g., [3], [4], [18]).
- 2) **Distributed processing:** In applications where an FC is absent, or where a large number of microphones signals need to be processed/transmitted simultaneously, it is desired to distribute the computational burden over the nodes of the WASN. In-network processing is often more energy efficient since the number of signals to be processed is small in each node². Furthermore, in-network processing can rely on nearest-neighbor transmissions, rather than long-distance transmissions to an FC.
- 3) **Bandwidth usage:** Because bandwidth is a scarce resource, it is important to use it as efficient as possible. If nodes only share data with their nearest neighbors (in a distributed setting), less transmission power is required and spatial reuse of the frequency spectrum is possible. Furthermore, to reduce the required communication bandwidth, compression of the data to be transmitted is of great importance to relax the requirements for the network layer. Compression and estimation are often jointly attacked in a WASN context, instead of treating them as independent problems (see Subsection IV-C).
- 4) **Scalability:** In the design of distributed algorithms for WASNs with many microphones, the goal is also to obtain a scalable algorithm in terms of communication bandwidth and/or local processing power. Basically, this means that adding an extra microphone has no (or limited) impact on the computational load or data traffic at the nodes that are not directly connected to this extra node. Distributed algorithms that allow simply connected networks are usually scalable [11], [14], [20], [21].
- 5) **Microphone subset selection:** In large-scale WASNs, sufficient performance can often be obtained by only using a subset of microphones (e.g., microphones that are close to a desired sound source). The less useful microphone nodes can then be put to sleep to save energy. The selection of a subset of useful microphones is a difficult problem on its own, and it is best tackled jointly with the estimation problem itself (see Subsection IV-D).
- 6) **Minimizing input-output delay:** The minimization of IO delay is an important challenge in real-time audio streaming WASNs, e.g., in hearing aids or telephony. An IO delay is introduced both at the DSP level [12] and at the network layer [22].
- 7) **Synchronization aspects:** Since each node of a WASN has its own clock, and since each clock's oscillator has imperfections, there is an inevitable clock skew³ and offset. Clock synchronization protocols and algorithms [23] are crucial for the data transmission in the communication layer, but also for multi-microphone audio processing algorithms, since their

²For most multi-microphone signal enhancement algorithms, the required computational power does not scale well with the number of microphone signals processed at a single device (e.g., quadratically [10], [19]).

³As a reference: the value for the (worst-case) clock skew, based on a 32kHz oscillator commonly used for sensor networks (i.a. in the Tmote Sky), is a difference of 40 ppm, i.e., approximately 40 μ s per second or 0.144 s in an hour [23], [24].

performance significantly degrades when the analog-to-digital converters (ADCs) of the different microphones sample at (slightly) different sampling rates [25]. In the case of signal enhancement, only clock skew has a negative effect on the performance, since this results in signal drift. A time-invariant clock offset is usually not that harmful; it is either inherently taken care of by the signal enhancement algorithm (e.g., in blind beamforming [3], [4], [18]–[20]), or it can be roughly estimated and compensated at start-up (e.g., based on cross-correlation techniques). However, clock offset may be harmful in other tasks, such as source localization.

In a WASN with dedicated and uniform hardware, synchronization of the sampling rates of ADCs is usually manageable [23] and sometimes even unnecessary if the oscillators are of sufficient quality⁴. On the other hand, in non-uniform ad-hoc WASNs with different devices from different manufacturers, synchronization of the ADCs may be hard (or impossible), and the resulting signal drift must then be taken into account by the signal processing algorithms.

Finally, it is noted that audio algorithms, that are not based on microphone signal coherence, can usually cope with significant ADC mismatch (e.g., energy-based methods [13], [26]).

- 8) **Routing and topology selection:** Intelligent routing decisions and topology selection are crucial in data-intensive WASNs. This is because of the strict timing requirements and the many different aspects that are involved in the decision making. The topology may be optimized in terms of transmission power, end-to-end delay, or QoS in general. Cross-layer interaction between the application layer and the network layer should ideally be incorporated in the decision making (more on this in Section V).

IV. DISTRIBUTED SIGNAL ENHANCEMENT FOR REAL-TIME AUDIO ACQUISITION WITH WASNs

One of the most difficult challenges for WASNs is real-time audio acquisition, including signal enhancement, e.g., for intelligible voice recording. By combining multiple microphone signals, an enhanced output signal can be obtained, where background noise is significantly reduced [1]. In these type of signal fusion applications lie the true challenges for the network layer design, since they produce a lot of data traffic, and they require reliable links with low packet loss, accurate synchronization, topology selection, small IO delay, etc. At the same time, there is a significant challenge on the signal processing level too, i.e., how to design suitable algorithms that relax these high-demanding requirements on the network layer design.

A. Suboptimal in-network fusion

There is a significant amount of literature on how to optimally fuse multiple microphone signals to exploit temporal and spatial correlation to reduce background noise, a.k.a. beamforming [1]. So-called 'blind' beamformers [1], [3], [4], [18], where the microphone and source positions are not assumed to be known, are particularly interesting for the WASN case, and can directly be applied if an FC is available (assuming that sufficient bandwidth is available, and that the clocks of the different microphones are synchronized). However, it is not obvious how these blind techniques can be applied to

⁴This only holds for *adaptive* audio processing algorithms. Furthermore, even if the application layer can handle (limited) mismatch in the ADC sampling frequencies, synchronization protocols are usually still required for the communication layer.

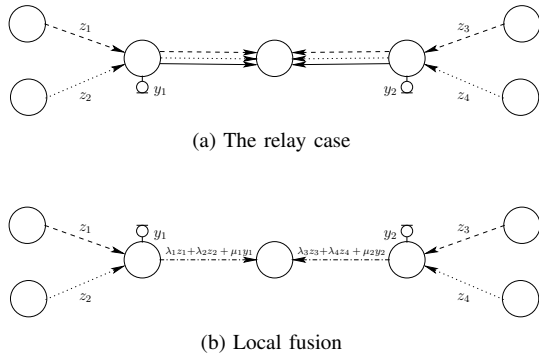


Fig. 6. Two different types of in-network data fusion.

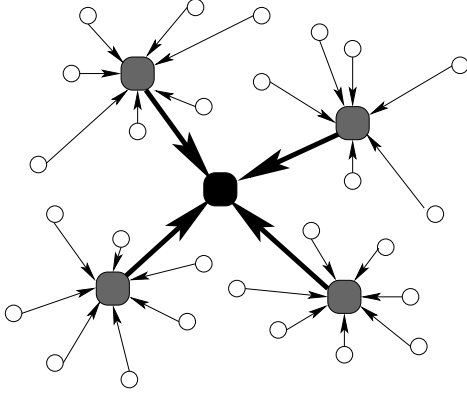


Fig. 7. Example of a node hierarchy for signal fusion.

decentralized topologies without a dedicated device that acts as an FC, i.e., where in-network processing is required.

A straightforward but naive approach to tackle this problem is to relay all the microphone signals to an arbitrary node in the network that then acts as an FC (see Fig. 6(a)). However, this does not scale well in terms of communication bandwidth and processing power, since the number of transmitted audio signals grows along the signal path, and the entire computational burden is borne by a single node. Furthermore, only in the node that was chosen as FC, an enhanced signal is available. A better approach is to let each node fuse its microphone signal(s) with the signal(s) obtained from neighboring node(s) into a single audio signal, and only transmit this fused signal to another node (see Fig. 5 and Fig. 6(b)). This approach obviously scales much better in terms of data traffic, and the computational burden is distributed over all the nodes in the network. Furthermore, many nodes then have access to an (at least partially) enhanced signal.

There are many ways to organize this distributed in-network fusion scheme. The most common approach is to construct a hierarchy of (local) centralized beamformers [11], [14], [21], i.e., the network is divided in clusters of nodes, and the nodes that form a cluster transmit their microphone signal(s) to a higher-level node, referred to as the cluster head (CH). The CH then basically has the role of a local FC for the nodes in the cluster, i.e., it fuses the received signals into a single enhanced audio signal. These locally enhanced signals at the different CHs are then fused together at the highest-level node (the data sink) to obtain the final output signal (see Fig. 7). This approach can easily be extended to deeper levels of hierarchy, which naturally leads to WASNs with a tree topology, where data flows from the lowest-level nodes (leaf nodes) to the highest-level node (root node),

fusing all the intermediate microphone signals.

This hierarchical architecture is elegant and perfectly scalable, but it has some major drawbacks. Even if every node is able to compute a locally optimal signal estimate based on the locally available signals (e.g., based on blind beamforming techniques [1], [3], [4], [18]), the final estimate at the data sink will still be suboptimal. For a given topology, computing the optimal set of fusion rules requires global information on the cross-correlation between all microphone signal pairs, which is usually not available⁵, especially in adaptive scenarios where these statistics must be estimated on the fly.

Furthermore, even if we would be able to compute the optimal set of fusion rules for a given topology, other topologies may exist that provide a better signal estimate. This results in a combinatorial problem, which is usually solved based on (suboptimal) heuristics [11], [14], [21].

B. Optimal in-network fusion

It is clear that the hierarchical fusion method described in the previous subsection is suboptimal due to the lack of global information, the many heuristics, and the dependence on the chosen topology. In particular in adaptive scenarios, the nodes must update their fusion rules based only on partial information, i.e., the local data to which they have access. However, there exist distributed adaptive speech enhancement techniques with in-network fusion that can generate an optimal signal estimate, independent of the chosen tree topology (assuming two-way data traffic between nodes) [20]. Furthermore, each node can enhance its own local microphone signal in an optimal way, as if all signals in the entire WASN were available to each node. This seems impossible, since information is inevitably thrown away after each fusion step. However, it can be shown that this is indeed possible in certain scenarios, i.e., where the number of desired speakers that need to be retained at the output is small [10], [12], [19], [20], [27], [28]. This class of algorithms is referred to as ‘distributed adaptive node-specific signal estimation’ (DANSE) algorithms. DANSE algorithms can operate in networks with a tree topology [20] or in a 2-level hierarchical network where the CHs (i.e., the grey nodes in Fig. 7) are fully connected with each other [12], [19].

The node-specific aspect of DANSE refers to the fact that each node may be interested in a different signal. For example, a binaural HA user wants to hear the sound as it impinges on his/her ears, and therefore the left HA will estimate a different signal than the right HA [8]–[10], [29]. This is important for directional hearing. Another example is sound source localization with a prior signal enhancement step to reduce noise in the recordings. In this case, the signal enhancement algorithm must preserve the node-specific target signal(s) as they are locally observed by the different microphone nodes.

The efficiency and optimality of DANSE relies on the assumption that the total number of desired speakers is much smaller than the number of available microphones. This is because the number of audio signals that each node needs to transmit is directly proportional to the number of desired source signals that need to be retained [10], [12], [19], [20], [28] (otherwise, optimality cannot be guaranteed). It is noted that, if the node-specific aspect of DANSE is relaxed, i.e., each node is interested in exactly the same signal, then the nodes only need to transmit a single audio signal, independent of the number of desired speakers.

⁵E.g., in Fig.6(b), the cross-correlation between the signals z_1 and z_4 cannot be computed since the nodes that generate them are not connected.

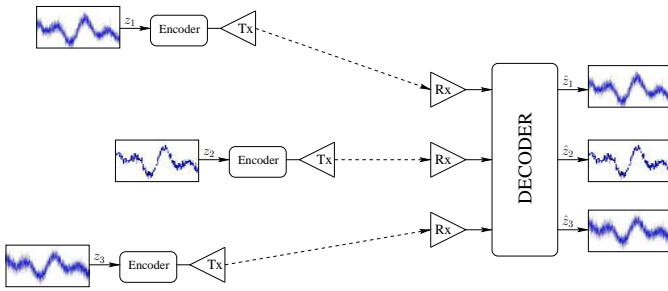


Fig. 8. A typical source coding scenario in a WASN: three nodes encode their locally preprocessed signals and transmit it to a fourth node, who decodes all three signals.

C. Distributed compression

The techniques addressed in the previous subsections aimed to estimate a signal in a distributed fashion, by fusing multiple audio channels into a single audio channel that is transmitted to neighboring nodes. To further reduce the bandwidth, and to match specific rate constraints in the wireless links, the transmitted signals must be encoded with efficient source coding techniques, and reconstructed at the the receiving node. The goal is then to transmit a signal as efficient as possible, without adding too much distortion between the original and the decoded signal.

A typical distributed source coding scenario is depicted in Fig. 8. The receiving node on the right collects encoded versions of three different signals (z_1 , z_2 and z_3) from three different nodes. The decoder at the receiving node needs to decode all three original signals and provide the reconstructed signals (\hat{z}_1 , \hat{z}_2 and \hat{z}_3) to the local fusion algorithm (e.g., DANSE). For node 1, the goal is thus to transmit the signal z_1 with the smallest possible distortion, given a certain available bit rate in the wireless link. Here, signal enhancement and coding can be viewed as cascaded techniques, but there also exist approaches where both are jointly tackled in a WASN context [30], [31]. It is obvious that such an integrated approach can yield better performance, since (lossy) compression inevitably has a negative effect on the signal enhancement algorithm (see, e.g., [32], which analyzes the effect of compression on DANSE).

The encoders in the distributed source coding scenario in Fig. 8 can be designed in two different ways. The simplest way is to merely encode the signal z_1 by removing the inherent redundancy in the signal z_1 itself. This is often referred to as ‘side information unaware’ (SIU) coding, since it ignores the mutual information in the signals of other nodes. However, since the receiving node also has access to encoded versions of z_2 and z_3 , and since these signals are usually highly correlated with z_1 , the latter can be transmitted with significantly less bits while keeping the same level of distortion. This is referred to as ‘side information aware’ (SIA) coding, i.e., the encoders are designed to jointly remove the mutual redundancy in all the signals z_1 , z_2 and z_3 [33], [34]. Obviously, SIA performs better than SIU, but an SIA encoder cannot be designed without prior knowledge on the mutual information in z_1 , z_2 and z_3 , which is often not available. However, in a WASN with 2-way data traffic (e.g., in DANSE-like algorithms), some SIA coding is possible, since a transmitting node then has access to (fused) data that is also available in the receiving node. Since the received signal and the signal to be transmitted will often have significant correlation, this can be exploited to significantly compress the transmitted signal. Furthermore, if a node transmits more than one audio signal to another node (as in DANSE with multiple desired speakers), the

transmitted signals can be jointly encoded by exploiting the cross-correlation between them.

D. Microphone subset selection

Another important aspect to facilitate the network layer design, and to relax the communication bandwidth and power constraints, is the selection of a subset of most useful microphone nodes. The other (less useful) nodes can then be switched off to reduce power consumption and data traffic in the WASN. An important question is on what basis this subset is selected. For example, in [35], the microphones with highest SNR are chosen, and in [36], a set of microphones is chosen that have a strong cross-correlation with each other, which is an important feature in the design of a beamformer. Other possible utility measures may be the direct-to-reverberant ratio, the microphone or speaker proximity, etc.

However, all these selection methods make abstraction of the signal enhancement algorithm that is used, which may be suboptimal. For example, a microphone that is close to an interfering source (e.g., a radio) has low SNR but may indeed be very useful for signal enhancement, i.e., as a noise reference to cancel this interferer in another microphone signal. Therefore, it is often advantageous to design the utility measure jointly with the signal enhancement algorithm, as in [37], [38]. Furthermore, the utility measures in [37], [38] can be computed efficiently from the available signal enhancement fusion rules at hardly any additional computational cost.

V. CROSS-LAYER DESIGN IN WASNS

In most WASNs, there is an important interaction between the application layer (DSP) and the network layer. Therefore, a joint design may significantly improve the performance of the WASN. A couple of the interactions between both layers are addressed below.

- The audio processing algorithm that is used often puts strict constraints on the topology of the network [11], [14], [19]–[21].
- The selected microphone subset also affects the choice of the topology since useless nodes (from a signal enhancement perspective) are removed from the network. Vice versa, the topology selection may also influence the subset selection algorithm, since certain nodes can be useful from a routing perspective. Furthermore, the ‘usefulness’ of a node also depends on the delay in the signal path from that node to other nodes (if the end-to-end delay is too large, the microphone signal may become useless in real-time applications).
- In large areas with long inter-microphone distances, many microphones will not be acoustically coupled due to the significant attenuation of sound over long distances. Such acoustical coupling can be easily detected by the audio processing algorithm (e.g. using cross-correlation techniques [36], or microphone utility [37], [38]). It is obvious that microphone nodes that are not acoustically coupled should not share data.
- The establishment of a wireless link between certain node pairs may require a large transmission power, e.g., due to shadow effects. However, this link may be very important from a signal enhancement perspective. This trade-off requires careful consideration. Another example is the design of a node hierarchy (see Subsection IV-A), which can be based on nearest neighbors to reduce transmission power, but this may yield suboptimal results in terms of signal enhancement performance.
- The network graph should depend on the quality of the microphone recordings at the different nodes. For example, a high-SNR node should ideally be positioned in the center of the network and/or close to data sinks, and it should have

many connections, such that this high-SNR signal can rapidly propagate through the network or to the end user, with a minimum number of hops.

VI. CONCLUSIONS

In this paper, we have addressed some possible applications that can benefit significantly from using WASNs, and we have listed the core challenges that need to be tackled in WASN design. We have given a general overview of distributed signal processing techniques for signal enhancement, and we have explained how these techniques can relax the high-demanding constraints on the network layer design. Finally, we have pointed out some interactions between the application layer and the network layer, which is a motivation for cross-layer design in WASN applications.

REFERENCES

- [1] M. Brandstein and D. Ward, *Microphone arrays: signal processing techniques and applications*. Berlin Heidelberg New York: Springer-Verlag.
- [2] H. Wang and P. Chu, "Voice source localization for automatic camera pointing system in videoconferencing," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, oct 1997.
- [3] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Transactions on Signal Processing*, vol. 50, no. 9, pp. 2230 – 2244, Sep. 2002.
- [4] 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, August 2009.
- [5] A. Bertrand, "Signal processing algorithms for wireless acoustic sensor networks," Ph.D. dissertation, Katholieke Universiteit Leuven, Leuven, Belgium, May 2011.
- [6] I. F. Akyildiz, T. Melodia, and K. R. Chowdhury, "A survey on wireless multimedia sensor networks," *Comput. Netw.*, vol. 51, pp. 921–960, March 2007.
- [7] H. Dillon, *Hearing Aids*. Boomerang Press, 2001.
- [8] B. Cornelis, S. Doclo, T. Van den Bogaert, M. Moonen, and J. Wouters, "Theoretical analysis of binaural multimicrophone noise reduction techniques," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 18, no. 2, pp. 342 –355, feb. 2010.
- [9] S. Doclo, T. van den Bogaert, M. Moonen, and J. Wouters, "Reduced-bandwidth and distributed MWF-based noise reduction algorithms for binaural hearing aids," *IEEE Trans. Audio, Speech and Language Processing*, vol. 17, no. 1, pp. 38–51, Jan. 2009.
- [10] 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. doi:10.1155/2009/530435.
- [11] Y. Jia, Y. Luo, Y. Lin, and I. Kozintsev, "Distributed microphone arrays for digital home and office," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, May 2006, pp. 1065–1068.
- [12] A. Bertrand, J. Callebaut, and M. Moonen, "Adaptive distributed noise reduction for speech enhancement in wireless acoustic sensor networks," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Tel Aviv, Israel, Aug. 2010.
- [13] M. Chen, Z. Liu, L.-W. He, P. Chou, and Z. Zhang, "Energy-based position estimation of microphones and speakers for ad hoc microphone arrays," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, oct. 2007, pp. 22 –25.
- [14] I. Himawan, I. McCowan, and S. Sridharan, "Clustered blind beamforming from ad-hoc microphone arrays," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 19, no. 4, pp. 661 –676, may 2011.
- [15] Y. Guo and M. Hazas, "Acoustic source localization of everyday sounds using wireless sensor networks," in *Proceedings of international conference adjunct papers on Ubiquitous computing*, ser. UbiComp '10. New York, NY, USA: ACM, 2010, pp. 411–412.
- [16] H. Wang, "Wireless sensor networks for acoustic monitoring," Ph.D. dissertation, University of California, Los Angeles (UCLA), Los Angeles, California, 2006.
- [17] E. Aarts and S. Marzano, *The New Everyday: Views on Ambient Intelligence*. 010 Publishers, 2003.
- [18] H. Buchner, R. Aichner, and W. Kellermann, "A generalization of blind source separation algorithms for convolutive mixtures based on second order statistics," *IEEE Trans. on Speech and Audio Processing*, vol. 13, no. 1, pp. 120–134, January 2005.
- [19] A. Bertrand and M. Moonen, "Distributed adaptive node-specific signal estimation in fully connected sensor networks – part I: sequential node updating," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5277–5291, 2010.
- [20] —, "Distributed adaptive estimation of node-specific signals in wireless sensor networks with a tree topology," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2196–2210, 2011.
- [21] Y. Hioka and W. B. Kleijn, "Distributed blind source separation with an application to audio signals," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, May 2011, pp. 233–236.
- [22] C. Wang, Y. Sun, and H. Ma, "Analysis of data delivery delay in acoustic sensor networks," in *IEEE/IFIP International Conference on Embedded and Ubiquitous Computing (EUC)*, vol. 1, dec. 2008, pp. 283 –287.
- [23] Y.-C. Wu, Q. Chaudhari, and E. Serpedin, "Clock synchronization of wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 124 –138, 2011.
- [24] M. J. Whelan and K. D. Janoyan, "Design of a robust, high-rate wireless sensor network for static and dynamic structural monitoring," *Journal of Intelligent Material Systems and Structures*, vol. 20, no. 7, pp. 849–864, 2009.
- [25] S. Wehr, I. Kozintsev, R. Lienhart, and W. Kellermann, "Synchronization of acoustic sensors for distributed ad-hoc audio networks and its use for blind source separation," in *IEEE International Symposium on Multimedia Software Engineering*, dec. 2004, pp. 18 – 25.
- [26] A. Bertrand and M. Moonen, "Energy-based multi-speaker voice activity detection with an ad hoc microphone array," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Dallas, Texas USA, March 2010, pp. 85–88.
- [27] —, "Distributed adaptive node-specific signal estimation in fully connected sensor networks – part II: simultaneous & asynchronous node updating," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5292–5306, 2010.
- [28] —, "Distributed node-specific LCMV beamforming in wireless sensor networks," *IEEE Transactions on Signal Processing*, 2011.
- [29] S. Markovich Golan, S. Gannot, and I. Cohen, "A reduced bandwidth binaural MVDR beamformer," in *Proc. of the Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Tel-Aviv, Israel, Aug. 2010.
- [30] O. Roy and M. Vetterli, "Rate-constrained collaborative noise reduction for wireless hearing aids," *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 645 –657, 2009.
- [31] J. Szurley, A. Bertrand, M. Moonen, P. Ruckebusch, and I. Moerman, "Utility based cross-layer collaboration for speech enhancement in wireless acoustic sensor networks," in *Proc. European signal processing conference (EUSIPCO)*, Barcelona - Spain, Aug. 2011, pp. 235–239.
- [32] T. C. Lawin-Ore and S. Doclo, "Analysis of rate constraints for MWF-based noise reduction in acoustic sensor networks," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011.
- [33] S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): design and construction," *IEEE Transactions on Information Theory*, vol. 49, no. 3, pp. 626–643, March 2003.
- [34] M. Gastpar, P. Dragotti, and M. Vetterli, "The distributed Karhunen-Loève transform," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5177 –5196, 2006.
- [35] M. Wölfel, C. Fügen, S. Ikbāl, and J. W. McDonough, "Multi-source far-distance microphone selection and combination for automatic transcription of lectures," in *Proc. INTERSPEECH*, 2006.
- [36] K. Kumatani, J. McDonough, J. Lehman, and B. Raj, "Channel selection based on multichannel cross-correlation coefficients for distant speech recognition," in *Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, 2011, pp. 1 –6.
- [37] 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, August 2010, pp. 1092–1096.
- [38] J. Szurley, A. Bertrand, and M. Moonen, "Efficient computation of microphone utility in a wireless acoustic sensor network with multi-channel Wiener filter based noise reduction," *Internal Report Katholieke Universiteit Leuven ESAT/SCD*, 2011.