MEMS microphones for wireless applications

8

J. Tiete*, F. Domínguez†, B. da Silva*, A. Touhafi*, K. Steenhaut*
*Vrije Universiteit Brussel, Elsene, Belgium, †Escuela Superior Politécnica del Litoral, Guayaquil, Ecuador

8.1 Introduction

Market research shows that the share of MEMS microphones in consumer products continues to grow over the previously ubiquitous electret condenser microphones (ECMs) [1]. An MEMS microphone is a miniature microphone, usually in the form of a surface mount device, that uses a miniature pressure-sensitive diaphragm as a way to pick up sound waves. This diaphragm is produced by surface micromachining of polysilicon on a silicon substrate or etched on a semiconductor using standard CMOS processes [2]. MEMS microphones are most commonly used in cellphones, hearing aids, tablets, laptop computers, video cameras, and more recently in Internet of Things (IoT) devices, wearables, medical devices, and cars (Fig. 8.1).

Fig. 8.1 MEMS microphone fabricated by Infineon Technologies AG.

Wireless MEMS Networks and Applications. http://dx.doi.org/10.1016/B978-0-08-100449-4.00008-7

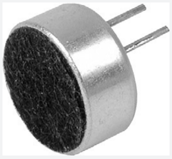
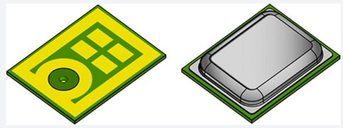
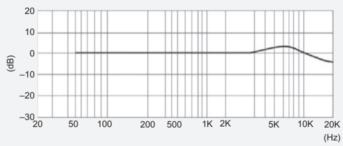
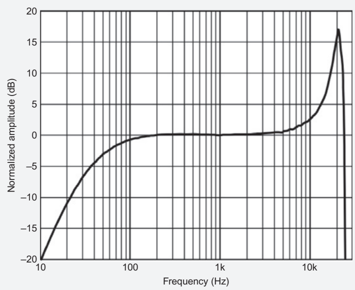
MEMS microphones have many advantages over ECMs and are therefore currently the preferred component choice in new electronic designs. The first obvious advantage is size, with die sizes of complete digital microphones being as small as 0.70 mm^2 (Akustica AKU230). MEMS microphones also have a higher performance density, with much better noise performance compared to an ECM of equal volume. They have a lower vibration sensitivity, due to the very low surface density of the membrane, as well as better temperature stability. The sensitivity of an MEMS microphone may only drift ± 0.5 dB over its operating temperature, while an ECM may drift as much as 4 dB over the same range. ECMs cannot be reflow soldered because the charge on an ECM's diaphragm cannot withstand the high temperatures during the reflow soldering process. MEMS microphones, on the other hand, can be mounted and reflowed in the same process as most other components on a printed circuit board (PCB). Having a more uniform part-to-part frequency response than ECMs makes MEMS microphones more suitable for applications where matched microphones are needed, such as microphone arrays (Table 8.1).

MEMS microphones can have a simple analog output, like a traditional ECM, but they can also have a digital output. These digital MEMS microphones output their audio signal in a serial data stream like I^2S or the more often used pulse density modulated (PDM) digital audio stream. Since MEMS microphones are produced on a silicon substrate, additional circuitry can easily be added on the same die or integrated in the small package of the microphone. This allows the analog-to-digital converters needed for the implementation of an analog microphone in a digital system to be implemented in the microphone itself. As a result, there is no analog signal chain on the PCB and fewer precautions need to be taken to combat electromagnetic interference (EMI) in the electronic design. Using digital MEMS microphones thus results in a lower overall component count and a smaller footprint, which lowers cost and design complexity. The design also becomes more flexible because a change in microphone supplier wouldn't mean a redesign of the complete analog signal chain.

Because of the reduced amount of components and resulting low power consumption (mW range for a single digital MEMS microphone) and better immunity against EMI originating from an antenna or capacitive touchscreen, MEMS microphones become the ideal solution for battery powered wireless devices.

In this chapter we will explore some applications of MEMS microphones used in wireless battery powered devices. In particular we will examine the use of MEMS microphones in cell phones, tablets, and laptop computers and why the implementation of an array of multiple microphones recently became possible and is beneficial for the user. We will introduce miniature microphone arrays and the workings of microphone arrays and their applications in general. In Section 8.4 we will introduce wireless sensor networks (WSNs) and how, with the aid of MEMS microphones and microphone arrays, they can be used for audio monitoring. Section 8.5 goes deeper into a specific application of an acoustic WSN and explains the processing needed to use a miniature microphone array and a WSN for sound source localization. We finally present our conclusions in Section 8.6.

Table 8.1 Comparison between a popular MEMS microphone and ECM (see data sheets for more detailed info)

	ECM	MEMS
Part number	CMA-4544PF-W	INMP621
Image		
Directionality	Omnidirectional	Omnidirectional
Output type	Analog	Digital (PDM)
Sensitivity	−44 dBV	−46 dBV
Signal-to-noise ratio	60 dBA	65 dBA
Size	9.7 mm Dia × 4.7 mm H	4 × 3 × 1 mm
Frequency range	20 Hz–20 kHz	45 Hz–20 kHz
Frequency response		
Price (1 K volume)	$0.405	$2.256

8.2 MEMS microphones for mobile applications

Nowadays, MEMS microphones are most commonly used in smartphones. The first cell phones that used MEMS microphones were introduced in 2003 and one of the big milestones for the small devices was when Apple started using them, with the introduction of the iPhone 4 in 2010 [3]. Currently, almost all smartphone manufacturers replaced their ECMs with MEMS microphones. Most smartphones use two or more microphones; one is usually placed on the front of the device to pick up the user's voice, and the other microphone is placed on the back of the device to pick up background noise. With this setup the device can perform background noise cancelation, resulting in a clearer recorded voice signal. The extra microphones also result in higher quality audio when recording video with the device's back- or front-facing camera (Fig. 8.2).

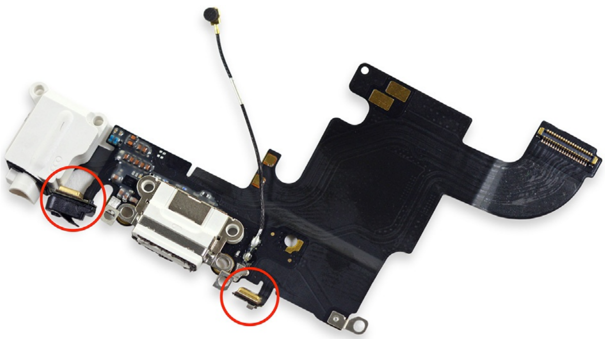


Fig. 8.2 iPhone 6s lightning cable assembly, with two MEMS microphones circled in red.

MEMS microphones have many advantages over ECMs when used in mobile applications. The small size permits manufacturers to not only implement multiple microphones in the device itself, but also in the accompanying headphones. Current headphones already have one MEMS microphone implemented to pick up the user's voice, but future models might implement multiple microphones on the wire leading to the earbuds and in the earbuds themselves, like the model described in a recent patent filing from Apple [4].

MEMS microphones in smartphones and tablets are designed to pick up audio from the user's voice when making a phone call and the surroundings when recording video. The quality of these audio recordings is increasingly improving. Some manufacturers even use this superior audio recording quality as their main selling point for their products. These high quality audio recordings can also be used by different applications running on the smartphone or tablet [5], for example, evaluated different sound measurement applications on different smartphones and compared the reported values with a high quality sound level meter. They concluded that smartphones can be used for noise pollution monitoring since they give a good representation of the current sound pressure level of the surroundings. This enables participatory noise pollution monitoring studies to use mobile phones [6].

8.3 Microphone arrays

Microphone arrays come in different shapes and sizes. They date back over 100 years and were first used by the military to determine the bearing and location of aircrafts, ships, and submarines. They consist of multiple microphones positioned in such a way that the spatial acoustic information can be properly captured. The fundamental theory used to process the signals produced by these sensor arrays is based on wave propagation. A microphone array can capture a sound signal from several different points simultaneously, which allows, with the proper processing, for spatial audio filtering. This means that with a microphone array, one can choose a point in space and filter out only the sound waves originating from that direction (Fig. 8.3).

Fig. 8.3 The Norsonic Nor848A 0.4 m diameter acoustic camera, which uses a 128 MEMS microphone array to graphically represent the surrounding soundscape.

Microphone arrays are currently being used in many different industries. The military industry is using microphone arrays to determine the flight path of bullets in order to determine the location of a shooter [7]. The automobile industry uses microphone arrays to identify the source of wind noise and rattling components on and in a car during the design phase. Many industries where noise can become a problem or is a precursor for failure, like wind turbines, rolling stock, airplanes, or appliances with electric motors or rotating parts in them, can benefit from a microphone array to pinpoint the exact sources of problems. In many cases where a microphone array is being used to locate a certain noise source, it will be complemented with a visual camera to form an acoustic camera. A heat map of the measured sound field by the microphone array is then overlaid on the video captured by the camera. This forms a video clip that visually represents the sound power emitting from all the objects in the environment.

The shape and size of the array as well as the positioning and number of microphones used in the array are closely linked to the performance of the microphone array. Depending on which signal processing technique is used to process the audio signals coming from the microphones in the array, the distance between microphone pairs and the overall size of the array will determine the resolution at which an array can image a certain frequency of sound incident. Delay-and-sum beamforming, for example, is a signal processing technique that enables directional signal or audio reception based on the phase difference of the received signals. With this technique, a smaller distance between microphones in the array will benefit the reception of higher frequency audio sources, and a bigger distance between microphones will benefit lower frequency signals. Microphone arrays also do not have to be laid out in only two dimensions. A spherical microphone array can, for example, be used to form a 3-D visual soundscape of the inside of a car or plane. The positioning of the microphones in a spherical configuration will enable the direction, or beam, that is being received to be anywhere in space (Fig. 8.4).

Fig. 8.4 A spherical microphone array by Bruël and Kjær, measuring the soundscape inside a car.

An array can also vary heavily in size, with the larger arrays spanning more than 100 m; they can be used, for example, for detecting the seismic activity of a volcano [8]. However some arrays are composed of MEMS microphones that are only several centimeters in diameter. These smaller arrays can, for example, be used in conference rooms to detect which person is currently speaking or in laptop computers to only record the sound coming from the user in front of the computer and attenuate all background noise.

8.3.1 Miniature MEMS microphone arrays

Because of the recent miniaturization and price drop of good quality microphones in the form of MEMS microphones, a new class of microphone arrays is rapidly developing. Previously, high-quality microphone arrays were comprised of multiple, expensive—several thousands of euros—and omnidirectional microphones with a perfectly flat frequency response. Every microphone in these arrays also needed a slew of high quality processing tools like phantom power supplies, preamplifiers, high quality cables, and analog-to-digital converters to make the array function. All this made microphone arrays very expensive and labor intensive to set up and use.

MEMS microphones are cheap, (less than one euro in volume) small, and have a decent signal-to-noise ratio (SNR) and frequency response. These features enable the production of miniature arrays that are several centimeters in diameter, on which the same signal processing techniques can be applied as their bigger counterparts. These miniature arrays can easily be mass produced, since a PCB can be produced in the desired array shape and the microphones can be mounted with standard PCB manufacturing techniques. Many consumer products like laptop computers, cell phones, and video cameras are already outfitted with more than one MEMS microphone and perform basic background noise cancelation with this limited array. Some laptop computers even use a linear array of several microphones along the top of the screen to only pick up speech coming from the user by using beamforming techniques.

Miniature MEMS microphone-arrays enabled the production of complex, small, stand-alone, and battery powered acoustic sensors. Processing of the digital audio signal coming from a miniature microphone array can easily be accomplished by a dedicated digital signal processor (DSP) (eg, STA321MP from STMicroelectronics) or a small FPGA. For example, Zhang et al. [9] designed several small—multiple centimeters in diameter—MEMS microphone arrays in order to identify and detect the direction of outdoor moving vehicles.

These cheap miniature arrays can easily be produced in volume and combined with a WSN to form an acoustic noise WSN that can, because of its distributed nature, better locate noise sources [10].

8.4 Acoustic noise WSNs with MEMS microphones

Acoustic noise WSNs measure sound in their surrounding environments to produce a soundscape with relatively high spatial and temporal resolution. They provide environmental data with high spatio-temporal resolution at the expense of data quality using cheap sensor nodes in large numbers. MEMS microphones are usually the acoustic sensor of choice in WSN implementations due to their low cost, small form factor, and relatively low power consumption. These characteristics are critical in battery-operated sensor nodes where unit costs and power consumption must be kept low.

8.4.1 WSNs

A WSN consists of a group of wirelessly interconnected, spatially distributed computers using sensors and actuators to assess and interact with their surrounding environment. These computers, known as sensor nodes, are usually resource-constrained, small-form-factor devices designed to run on battery power. Typically, they communicate with a remote gateway via an uplink connection to a network backbone or via their closest neighbor in a wireless mesh network (Fig. 8.5).

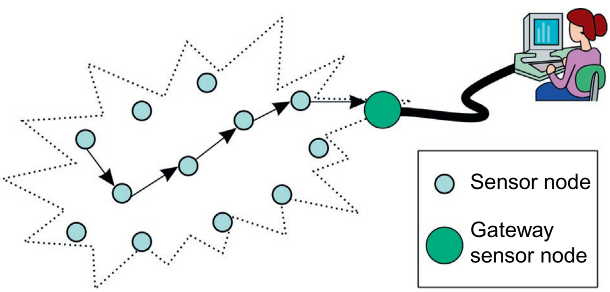


Fig. 8.5 Sensor nodes in a wireless sensor network (WSN) measure environmental parameters and transmit the data back to a network gateway where it is aggregated, stored, and processed.

WSNs were first proposed for military applications by the Department of Defense of the United States as a tool to detect and track enemy troops. Nowadays, they are applied in several application domains such as the IoT, indoor tracking, environmental monitoring, home automation, and industrial automation [11,12].

Sensor nodes are the main components of a WSN, and their functionalities are [13] as follows: data acquisition from different sensors; buffering and caching of sensor data; data processing; self-testing and monitoring; reception, transmission, and forwarding of data packets; and coordination of networking tasks. Sensor nodes are usually battery operated and deployed in large numbers in difficult to reach places. These constraints steer the design of sensor nodes towards maximum power efficiency and minimum unit cost.

Sensor nodes, depending on their resource usage, can be divided in three broad categories [13]:

- *SBC* (*Single-Board Computer*): These nodes have enough computing power to run full-fledged operating systems such as Linux or Windows. Their power consumption ranges from a few hundred milliwatts (eg, Raspberry Pi) to several watts (eg, ALIX 3D3 from PC Engines). While deployed on-the-field, they rely on external energy sources such as solar power or mains.
- *High-end ESM* (*Embedded Sensor Module*): These nodes are based on system-on-chip (SoC) microcontrollers where the CPU, RAM, flash, and several other peripherals (ADC, DAC, GPIOs, RF, etc.) are embedded in a single die. They have enough computer resources to run embedded versions of well-known operating systems. These devices typically consume several hundred milliwatts of power (eg, Intel Edison) and are typically expected to run on batteries for short periods (weeks, months).

- *Low-end ESM*: These nodes are resource constrained devices (a few kilobytes of RAM, CPU running typically at frequencies below 100 MHz) designed to run on batteries for long periods (months, years). They are based on low-power, low-cost SoC microcontrollers (eg, TI CC2538, Freescale MC13234) that directly implement power efficient RF communication stacks such as the IEEE 802.15.4 [14]. They require resource-optimized operating systems such as TinyOS or Contiki and are configured to the network's specific application. Their power consumption is typically in the order of a few milliwatts.

Low-end and high-end ESMs are usually found in IoT applications and high spatio-temporal resolution WSNs, while SBCs nodes usually serve as network sinks or as data collection nodes where intensive data processing is required on-site (Fig. 8.6).

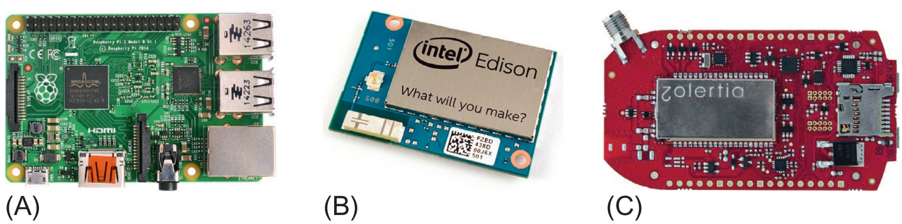


Fig. 8.6 Sensor nodes can be divided into three categories: single-board computers (SBC) (A), high-end ESM (embedded sensor module) (B), and low-end ESM (C).

8.4.2 Audio monitoring

In the context of this chapter, audio monitoring is the task of collecting and recording sounds from the surrounding environment with the goal of understanding their origins, processes, and effects. These sounds may either be anthropogenic, such as acoustic noise pollution in urban or industrial environments, or naturogenic, such as sounds originating from volcanic activity or wildlife.

Audio monitoring is a critical component of acoustic noise pollution studies in urban environments. Noise pollution is a well-known human stressor and has been linked to health problems such as cardiovascular disease, stress, and sleep deprivation [15,16]. Audio monitoring in urban areas helps to determine the where, when, how much, and who of noise pollution by creating high spatio-temporal resolution soundmaps. Examples of such studies are the RUMEUR network in France [17], Sensor City in The Netherlands [18], and IDEA in Belgium [19]. Additionally, audio monitoring has been used as a proxy measurement for various sources of air pollution [20].

On the naturogenic side, audio monitoring contributes to the tasks of ecosystem monitoring and wildlife observation [11,21]. Audio monitoring can be used to infer the presence, state, and movements of wildlife in a geographical area. However, due to the wide audio spectrum of wildlife, naturogenic audio also encompasses infrasound (elephants) and ultrasound (bats, dolphins) [21]. In this domain, audio monitoring is typically used to collect geotagged sounds produced by wildlife in its natural habitat to be later, by a human expert or a machine learning technique, curated and classified. Its main advantage in ecosystem monitoring and wildlife observation is its non-intrusiveness, while its main disadvantage is the requirement of high spatial resolution and resource-heavy data processing.

8.4.3 WSNs with MEMS microphones

Traditionally, applications and scientific studies that required geographical audio monitoring deployed trained personnel or volunteers in situ to manually collect audio data [22]. WSNs significantly automate and improve this process due to their capacity to automatically collect and distribute data in real time with high spatio-temporal resolution.

To provide reliable, high quality sound measurements, acoustic WSNs usually relied on professional measurement microphones to collect audio data [17,18]. These microphones are typically calibrated in a certified acoustic laboratory and exhibit a relatively high dynamic range and SNR. Their price ranges from several hundred to several thousand euros.

Acoustic WSNs relying on professional measurement microphones have a high cost per sensor node, and consequentially, their collected sound measurements have a rather poor spatial resolution. However, for most audio monitoring applications, the dynamic range and SNR of low-cost electret and MEMS microphones is sufficient, and their relatively lower data quality and reliability could be compensated by their high numbers and high spatial resolution [23,24].

MEMS microphones have several characteristics that are ideal for acoustic noise WSNs: compactness, low power consumption, and sufficiently high SNR and dynamic range [25,26]. Moreover, these characteristics have been improving in the last few years, with commercially available MEMS microphones reaching SNR levels up to 65 dB [27]. These levels are comparable or better than most low-cost electret microphones.

MEMS microphones can be divided in two categories: analog and digital. Analog MEMS are usually smaller and consume less power than their digital counterparts but require the addition of an analog-to-digital converter (ADC) plus other components of the audio measurement chain. Digital MEMS microphones incorporate several audio measurement chain components (ADC, preamp, signal conditioning filters, etc.) within the same enclosure and output sound data either using I^2S or PDM. Considering the extra resources required for an external ADC and signal conditioning, digital MEMS microphones are more power and space efficient than analog ones [25].

Tan and Jarvis [26] used an analog MEMS microphone, ADMP401, to build an energy harvesting acoustic sensor node. The main motivation was the extremely low power consumption of the sensor, below 250 μA [10]. selected a digital MEMS microphone, the ADMP521, to build a 52-microphone array on-board a sensor node in a WSN. The main motivation was the savings in power (the microphone consumes 900 μA while active and 1 μA in sleep mode), computing resources, and space.

8.5 Sound source localization with WSNs and MEMS arrays

8.5.1 Microphone array processing

Many MEMS-based applications such as sound-source localization need devices composed of multiple MEMS microphones. This leads to a high number of inputs and outputs (I/O) that need to be processed in real-time. Standard microcontrollers and DSPs offer a limited number of I/O. Furthermore, the complexity of the signal processing system drastically increases with the amount of MEMS involved, which becomes too computationally intensive for traditional signal processing devices. The scalability and processing problems can be solved using programmable logic devices such as a field programmable gate array (FPGA). FPGAs have a high number of I/O available and offer massive parallel computational power, which simplifies and accelerates the signal processing. Consequently, FPGAs are often considered to process the output of analog or digital MEMS microphone arrays.

8.5.1.1 What is an FPGA?

An FPGA is a semiconductor device composed of logic blocks interconnected via programmable connections. The logic blocks consist of look-up tables (LUTs) with a fixed number of inputs and are constructed over simple memories, SRAM or Flash, that store Boolean functions. Each LUT is coupled with a multiplexer and a flip-flop register in order to support sequential circuits. Likewise, several LUTs can be combined for implementing complex functions.

Today's FPGAs are powerful devices with support for dozens of I/O standards such as I^2C, SPI, CAN, PCIe. I/Os in FPGAs are grouped in banks where each bank is able to independently support different I/O standards.

FPGAs can be reprogrammed to the desired application or functionality requirements. The hardware description elaborated by the designer is used by the vendor's synthesizer in order to find an optimized arrangement of the FPGA's resources that implements the described functionality. This feature distinguishes FPGAs from application-specific integrated circuits, which are custom manufactured for specific design tasks.

Originally, FPGAs have been used in network packet analysis and signal processing. Thanks to high-speed embedded resources such as DSP slices and fast memories, FPGAs are now also utilized for algorithm acceleration either as coprocessors or standalone systems. In fact, theoretically, there is no limitation in the calculation speed of an FPGA since it performs parallel processing. Only the available FPGA resources determine the amount of operations that can be implemented.

8.5.1.2 FPGA-based systems using MEMS microphones arrays

As described in the previous section, digital MEMS microphones are commonly chosen to build microphone arrays. This kind of microphone moves the analog-to-digital conversion function from the DSP or the FPGA into the chip. Digital MEMS microphones integrate the transducer element together with an amplifier and an ADC. The encoded output format of digital MEMS microphones is often a PDM signal, which can be described as an oversampled 1-bit audio signal. This clocked digital signal is less prone to external noise than an analog signal and also enables two microphones, with the help of multiplexing, to use the same data and clock line. The protection from interfering signals makes digital MEMS microphones a preferred solution for handheld devices, for which analog audio signals may be susceptible to interference. Furthermore, the use of digital MEMS microphones reduces the complexity of the hardware since they do not require external amplifiers. However, the PDM signal needs to be demodulated to an analog form to be perceptible as audio, or it needs to be converted to multi-bit pulse code modulation (PCM) to be digitally analyzed Lewis [28]. A common strategy, detailed by Hegde [29], consists of multi-filter stages for the PDM demodulation and PCM conversion. The FPGA's capacity for highly parallel arithmetic architectures makes it well-suited for custom data path processes such as delay-and-sum beamforming algorithms.

One example of an MEMS microphone array exploiting FPGA's features is detailed by Salom et al. [30]. They propose a beamforming-based acoustic system for localization of the dominant noise source. The signal acquisition consists of a microphone array composed of up to 33 MEMS microphones, while the PDM demodulation and the beamforming is implemented in an FPGA. The authors implement the PDM demodulation described by Hegde [29]. The PDM demodulation starts with a PDM-to-PCM conversion by using cascaded integrator-comb (CIC) filters. This component is followed by a half-band low-pass decimation filter that reduces the sampling frequency. A final low-pass FIR filter is needed to remove the high-frequency noise introduced by the sigma-delta converter, which is integrated in the digital MEMS microphones. The implementation in the FPGA is completed with the delay-and-sum beamforming, measuring 60 angles, and a polar map is generated for directivity pattern presentation.

Another example is detailed by Perrodin et al. [31], where digital MEMS microphones are combined with FPGAs for robot-based applications. The authors propose an automated microphone array shape calibration to accurately estimate the array elements in order to face the noisy and reverberant environment of real-world robotic operations. Such calibration is based on time differences of arrival (TDOA) of moving sound sources. The FPGA performs the pre-processing stage proposed by Hegde [29], composed of a CIC filter and two half-band filters in order to downscale the sample rate, followed by a low-pass FIR filter to remove the high-frequency noise. The high amount of I/O pins available in the chosen FPGA allows for the connection of up to 128 digital MEMS microphones. Furthermore, thanks to the computational power of FPGAs, the pre-processing of the 128 digital MEMS microphones performed on the FPGA enables real-time operations.

Despite the fact that the authors in Havránek et al. [32] do not provide detailed information, the authors propose a calibration methodology for microphone arrays using an FPGA. The FPGA however, is one more component of a large system used to validate their methodology.

Digital MEMS microphones have the potential to offer similar performance to high-quality analog microphones for some applications. A comparison between digital MEMS microphones and analog microphones is presented by Zwyssig et al. [33]. They describe the design and implementation of an eight-element digital MEMS microphone array for distant speech recognition, which is compared to an analog equivalent composed of eight high-quality analog microphones. The results show that the absolute difference in word-error-rate (WER) is around 14% worse for the digital array, but it can be reduced to 4.5% using recognition techniques.

An alternative approach is proposed by Sanchez-Hevia et al. [34], where the authors use an FPGA to implement an acoustic camera. Instead of applying traditional solutions using CIC filters to demodulate the PDM signals, the authors propose their own custom designed demodulator filter, a cascaded recursive-running sum filter. Their fine-tuned filter reduces the hardware complexity while offering the required filtering needs. As result, their FPGA-based acoustic camera generates real-time video representing the received sound pressure.

A digital MEMS microphone array is proposed by Tiete et al. [10] for sound-source localization. The authors propose a FPGA-based implementation with a different strategy to process the 52 microphones that comprise the array. Instead of implementing individual filters for each microphone, the authors propose the execution of the delay-and-sum beamforming algorithm over the PDM signals. The output of the delay-and-sum, which is no longer a PDM signal, is filtered by windowing and is processed at the frequency domain.

Fig. 8.7 depicts the different strategies for PDM signal processing of digital MEMS microphone arrays. There is a clear trade-off in terms of hardware resources when applying beamforming before or after the PDM demodulation. The 1-bit data storage of the PDM signals when beamforming before the PDM demodulation promises to be an interesting area and power saving technique. Because only one PDM demodulation stage is needed, only one filter would be required instead of one cascade of filters per microphone. However, the output of the delay-and-sum computation is not a pure PDM signal, and this might increase the filter stage complexity.

8.5.2 Audio data transport in WSNs

Most of the examples below use wired connections between the sensor arrays and the processing devices. However, this is not an option for WSNs. Several challenges need to be acknowledged when considering audio transmission over wireless protocols. The main challenges could be grouped by compression techniques, security, and power consumption of the communication.

Since WSNs are mainly composed by battery-based nodes, the power consumption becomes a crucial factor. Many real-time audio applications over sensor networks

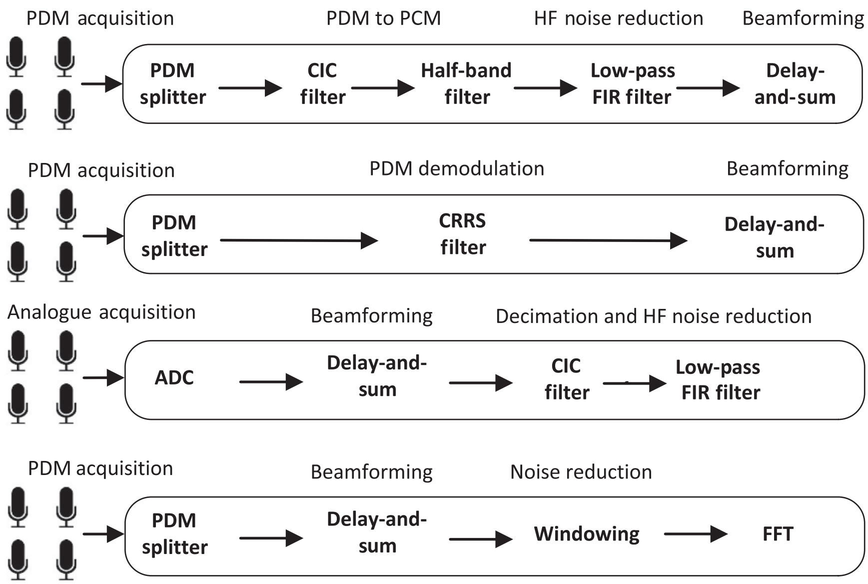


Fig. 8.7 Signal processing strategies for MEMS microphone arrays implemented on FGPAs.

have strong QoS requirements of delay and throughput. In such applications, the routing protocol has a predominant role in terms of energy [35]. An energy-efficient protocol not only needs to guarantee a reliable communication, but also low power consumption by minimizing the communication. Currently, several multipath routing tactics such as IEATH [36], QEMPAR [37] or CMQ [38] have been proposed for WSN for real-time applications. These multipath algorithms consider performance metrics of WSNs such as delay, energy consumption, and bandwidth while keeping the reliability of the communication system.

Several audio applications require security protection in order to avoid malicious intrusions while delivering audio data in WSNs. Audio watermarking is an effective authentication technique that embeds a small watermark into the original audio data. A quality-driven and energy-efficient watermarking system for audio transmissions in WSN is proposed in Wang et al. [39]. The authors propose embedding the watermark into the middle sub-bands, and it is robust against compression distortion. Such techniques reduce the power consumption and the data overhead, since the authentication protocol is embedded in the transmitted data.

Despite the existence of many audio compression techniques, only few of them target audio compression in WSN. An interesting distributed compression model for WSN [40] combines compressed and uncompressed audio data depending of the node's role. The compressed audio is decoded through correlations between the compressed and uncompressed data, achieving up to 40% of energy savings. Other techniques propose wavelet-based compressions [41,42]. An interesting power-efficient compression technique based on wavelet transformation is proposed by Dutta et al. [43]. Instead of the use of correlation, the authors proposed a dynamic difference detection technique, which eliminates the temporal correlation. Thus, instead of sending raw sample data, only the difference between two consecutive sample values is sent, reducing the data transmission cost. Such combination of wavelet transformation with difference detection reduces the energy consumption while preserving the decoding quality.

8.5.3 Data fusion

In a WSN that performs sound source localization, each sensor node collects only a portion of the information necessary to locate one or multiple sound sources. The information collected in each node must be transmitted and routed to a network gateway, where it will be aggregated or fused. This process, known as data fusion, integrates the partial data collected in each node and outputs an estimation of the geographical location of all detected sound sources. The data fusion algorithm depends on the overall localization technique (eg, triangulation) and is usually performed in a centralized high-performant network node.

As mentioned in the previous section, transmitting and collecting raw audio data is usually too resource-intensive for most sensor nodes equipped with microphone arrays. Consequently, each sensor node pre-processes the array's audio output and estimates the location or bearing of nearby sound sources. This estimation is then transmitted over the WSN to a network sink.

Astapov et al. [44] describe a WSN where each node has a two-microphone array. This array is used to estimate the direction of arrival (DOA) of a nearby sound source. The DOA, a real value representing an angle from −90 to 90 degrees, is transmitted over the network to a sink. The position of each possible sound source is then triangulated based on all collected DOAs and the fixed location of each node. The authors do not specify the accuracy of this method in this particular scenario but suggest using it to constrain the sound source search area for more resource-intensive microphone array localization algorithms such as steered response power with Phase transform (SRP-PHAT) [45].

Tiete et al. [10] use a 52-microphone array in each node to estimate the bearing of nearby sound sources. In each node, the array beam is steered in a 360 degrees sweep and the power at each angle in the sweep is measured. The result is a tuple of n power measurements, where n is the number of angles used to perform the sweep–the optimal value of n was determined to be 64 angles. The tuple, also known as the polar steered response power (P-SRP), is transmitted over the network to a sink.

At the sink, the P-SRP, together with the node's known position, is used to generate a probability map of the location of detected sound sources (Fig. 8.8). The probability map generated by each node is a matrix with elements ranging from 0 to 1. The matrix represents the geographical area covered by the WSN, and each element in the matrix represents the probability that a sound source is located at that point. The fusion algorithm simply adds the probability map generated by each node and finds all local

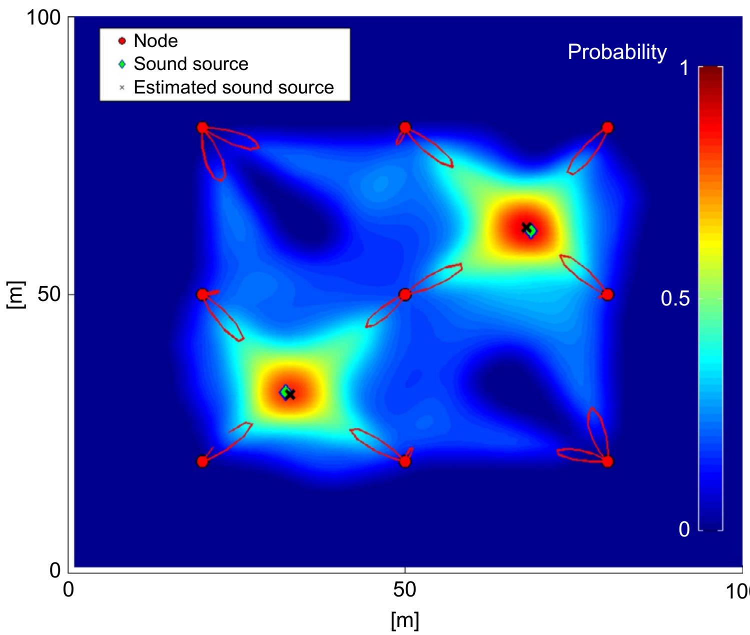


Fig. 8.8 Each WSN node uses an on-board microphone array to estimate the direction of nearby sound sources *(red lines)*. These estimations are then combined to form a probability map of the possible geographic location of all sound sources.

maxima in the resulting aggregated map. The local maxima represent the location of detected sound sources. The authors reported an experimental location accuracy below 10 cm in a 25 m^2 area.

8.6 Conclusion

MEMS microphones are ideal for wireless applications such as mobile phones, audio monitoring WSNs, and sound source localization WSNs. These applications typically require low power consumption, miniature sizes, and relatively high data quality. MEMS microphones now provide SNRs of up to 65 dB, with a PCB footprint from 8 to 12 mm^2. They consume very little power, in the range of a few milliwatts, and some digital versions support sleep mode, where power consumption drops to a few microwatts. Digital MEMS microphones include in one package almost the entire audio measurement chain, providing high area and power efficiency at a low cost (below 1 euro for some models). These characteristics will most likely keep improving, enabling the adoption of audio capturing and processing in IoT and WSN applications.

References

[1] M. Boustany, J. Bouchaud, MEMS microphones report 2014, in: IHS iSuppli, Technical Report, 2014.

[2] J.J. Jr Neumann, K.J. Gabriel, A fully-integrated CMOS-MEMS audio microphone. Digest Tech. Papers 1 (412) (2003) 230–233, http://dx.doi.org/10.1109/SENSOR.2003.1215295.

[3] R. Bogue, Recent developments in MEMS sensors: a review of applications, markets and technologies. Sens. Rev. 33 (4) (2013) 300–304, http://dx.doi.org/10.1108/SR-05-2013-678.

[4] S.V. Dusan, A. Lindahl, System and Method of Mixing Accelerometer and Microphone Signals to Improve Voice Quality in a Mobile Device. US Patent App. 13/840,667; 2014. https://www.google.com/patents/US20140270231.

[5] C.A.K. Kardous, P.B. Shaw, Evaluation of smartphone sound measurement applications. J. Acoust. Soc. Am. 135 (4) (2014) EL186–EL192, http://dx.doi.org/10.1121/1.4865269.

[6] S. Walkowiak, L. Ligowski, K. Wawruch, Information technologies in environmental engineering. information technologies in environmental engineering. Environ. Sci. Eng. 3 (3) (2011) 749–762, http://dx.doi.org/10.1007/978-3-642-19536-5.

[7] J. Sallai, W. Hedgecock, P. Volgyesi, A. Nadas, G. Balogh, A. Ledeczi, Weapon classification and shooter localization using distributed multichannel acoustic sensors, J. Syst. Archit. 57 (10) (2011) 869–885, http://dx.doi.org/10.1016/j.sysarc.2011.04.003. URL. http://linkinghub.elsevier.com/retrieve/pii/S138376211100052X.

[8] R.S. Matoza, M.A.H. Hedlin, M.A. Garcés, An infrasound array study of Mount St. Helens. J. Volcanol. Geotherm. Res. 160 (3-4) (2007) 249–262, http://dx.doi.org/10.1016/j.jvolgeores.2006.10.006.

[9] X. Zhang, J. Huang, E. Song, H. Liu, B. Li, X. Yuan, Design of small MEMS microphone array systems for direction finding of outdoors moving vehicles, Sensors (Switzerland) 14 (3) (2014) 4384–4398, http://dx.doi.org/10.3390/s140304384.

[10] J. Tiete, F. Domínguez, B. da Silva, L. Segers, K. Steenhaut, A. Touhafi, SoundCompass: a distributed MEMS microphone array-based sensor for sound source localization, Sensors 14 (2) (2014) 1918–1949. http://www.mdpi.com/1424-8220/14/2/1918.

[11] A. De La Piedra, F. Benitez-capistros, F. Dominguez, A. Touhafi, Wireless sensor networks for environmental research: a survey on limitations and challenges, in: EuroCon, Zagreb, Croatia, IEEE, New York, 2013, pp. 267–274.

[12] P. Rawat, K.D. Singh, H. Chaouchi, J.M. Bonnin, Wireless sensor networks: a survey on recent developments and potential synergies, J. Supercomput. 68 (1) (April 2014) 1–48.

[13] L.M. Oliveira, J.J. Rodrigues, Wireless sensor networks: a survey on environmental monitoring, J. Commun. 6 (2) (2011) 143–151.

[14] J. Weiss, New Low-power Wireless Standards Expand Wireless Sensor Network Applications, in: Wireless Congress 2011: Systems & Applications, Munich, Germany. Dust Networks, 2011.

[15] H.W. Davies, J.J. Vlaanderen, S.B. Henderson, M. Brauer, Correlation between co-exposures to noise and air pollution from traffic sources, Occup. Environ. Med. 66 (5) (2009) 347–350.

[16] A.V. Moudon, Real noise from the urban environment: how ambient community noise affects health and what can be done about it, Am. J. Prev. Med. 37 (2) (2009) 167–171.

[17] C. Mietlicki, F. Mietlicki, M. Sineau, An innovative approach for long-term environmental noise measurement: RUMEUR network, in: Institute of Noise Control Engineering (Ed.), INTER-NOISE and NOISE-CON Congress and Conference Proceedings, INCE, New York, NY, 2012, pp. 7119–7130.

[18] K. Friso, J. Reitsma, Sensor city Assen: Onderweg met op maat gesneden verkeersinformatie, in: Colloquium Vervoersplanologisch Speurwerk, CVS, Antwerpen, Belgium, 2011, pp. 1–11.
[19] D. Botteldooren, T.V. Renterghem, D. Oldoni, D. Samuel, L. Dekoninck, P. Thomas, W. Wei, M. Boes, B. De Coensel, B. De Baets, B. Dhoedt, The internet of sound observatories, in: Proceedings of Meetings on Acoustics, Montreal, Canada, vol. 19, Acoustical Society of America, Melville, NY, 2013, pp. 1–7.
[20] L. Dekoninck, D. Botteldooren, L.I. Panis, S. Hankey, G. Jain, S. Karthik, J. Marshall, Applicability of a noise-based model to estimate in-traffic exposure to black carbon and particle number concentrations in different cultures, Environ. Int. 74 (2015) 89–98.
[21] M. Baratchi, N. Meratnia, P.J.M. Havinga, A.K. Skidmore, B.A.G. Toxopeus, Sensing solutions for collecting spatio-temporal data for wildlife monitoring applications: a review, Sensors (Basel, Switzerland) 13 (5) (2013) 6054–6088.
[22] T.L. McDonald, Review of environmental monitoring methods: survey designs, Environ. Monit. Assess. 85 (3) (July 2003) 277–292.
[23] S. Dauwe, D. Oldoni, B. De Baets, T. Van Renterghem, D. Botteldooren, B. Dhoedt, Multi-criteria anomaly detection in urban noise sensor networks, Environ. Sci. Process. Impacts 16 (2014) 2249–2258.
[24] T. Van Renterghem, P. Thomas, F. Dominguez, S. Dauwe, A. Touhafi, B. Dhoedt, D. Botteldooren, On the ability of consumer electronics microphones for environmental noise monitoring, J. Environ. Monit. 13 (3) (March 2011) 544–552.
[25] J. Lewis. MS-2472: Analog and Digital MEMS Microphone Design Considerations. Technical Report, Analog Devices, 2013.
[26] W.M. Tan, S.A. Jarvis, On the design of an energy-harvesting noise-sensing WSN mote, EURASIP J. Wirel. Commun. Netw. 2014 (167) (2014) 1–18.
[27] J. Lewis, P. Schreier. Low Self Noise: The First Step to High Performance MEMS Microphone Applications. Technical Report, InvenSense, 2013.
[28] J. Lewis, Understanding microphone sensitivity, Analog Dialog. 46 (2012).
[29] N. Hegde, Seamlessly interfacing MEMS microphones with blackfin processors, in: EE-350 Engineer-to-Engineer Note, 2010.
[30] I. Salom, V. Celebic, M. Milanovic, D. Todorovic, J. Prezelj, An implementation of beamforming algorithm on FPGA Platform with digital microphone array, in: Audio Engineering Society Convention 138, Audio Engineering Society, New York, 2015.
[31] F. Perrodin, J. Nikolic, J. Busset, R. Siegwart, Design and calibration of large microphone arrays for robotic applications, in: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2012, pp. 4596–4601. http://dx.doi.org/10.1109/IROS.2012.6385985.
[32] Z. Havránek, P. Beneš, S. Klusáček, Free-field calibration of MEMS microphone array used for acoustic holography, in: Proceedings of the 21st International Congress on Sound and Vibration, 2014.
[33] E. Zwyssig, M. Lincoln, S. Renals, A digital microphone array for distant speech recognition, in: 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), IEEE, New York, 2010, pp. 5106–5109.
[34] H. Sanchez-Hevia, R. Gil-Pita, M. Rosa-Zurera, et al., FPGA-based real-time acoustic camera using PDM MEMS microphones with a custom demodulation filter, in: IEEE 8th Sensor Array and Multichannel Signal Processing Workshop (SAM), 2014, pp. 181–184.
[35] P. Jayanthi, T. Parameswaran, A survey on energy efficient multipath routing in wireless sensor network, Int. J. Adv. Inf. Commun. Technol. 1 (8) (2014) 730–735.

[36] S. Saqaeeyan, M. Roshanzadeh, IEATH: improved energy aware and two hop multipath routing protocol in wireless sensor networks, Int. J. Comput. Netw. Inf. Sec. 4 (5) (2012) 22–28.
[37] S.R. Heikalabad, H. Rasouli, F. Nematy, N. Rahmani, QEMPAR: QoS and energy aware multi-path routing algorithm for real-time applications in wireless sensor networks, Int. J. Comput. Sci. 8 (2011) 466–471.
[38] A.R. Rezaie, M. Mirnia, CMQ: clustering based multipath routing algorithm to improving QoS in wireless sensor networks, Int. J. Comput. Sci. 9 (3) (2012) 156–160.
[39] H. Wang, W. Wang, M. Chen, X. Yao, Quality-driven secure audio transmissions in wireless multimedia sensor networks, Multim. Tools Appl. 67 (1) (2013) 119–135.
[40] J. Chou, D. Petrovic, K. Ramchandran, Tracking and exploiting correlations in dense sensor networks, in: Conference Record of the Thirty-Sixth Asilomar Conference on Signals, Systems and Computers, vol. 1, IEEE, New York, 2002, pp. 39–43.
[41] A. Ciancio, A. Ortega, A distributed wavelet compression algorithm for wireless sensor networks using lifting, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004. Proceedings (ICASSP'04), vol. 4, IEEE, New York, 2004.
[42] G. Luo, Ultra low delay wavelet audio coding with low complexity for real time wireless transmission, in: Proceedings of 2005 International Symposium on Intelligent Signal Processing and Communication Systems, 2005. ISPACS 2005, IEEE, New York, 2005, pp. 741–744.
[43] I. Dutta, R. Banerjee, T. Acharya, S. DasBit, An energy efficient audio compression scheme using wavelet with dynamic difference detection technique in wireless sensor network, in: Proceedings of the International Conference on Advances in Computing, Communications and Informatics, ACM, New York, 2012, pp. 360–366.
[44] S. Astapov, J.-S. Preden, J. Ehala, A. Riid, Object detection for military surveillance using distributed multimodal smart sensors, in: 19th International Conference on Digital Signal Processing, Hong Kong, China, IEEE, New York, 2014, pp. 366–371.
[45] J.H. DiBiase, A High-Accuracy, Low-Latency Technique for Talker Localization in Reverberant Environments Using Microphone Arrays, Ph.D. Thesis. Brown University, Providence, RI, 2000.