

## Article

# A Low Cost Wireless Acoustic Sensor for Ambient Assisted Living Systems

Miguel A. Quintana-Suárez <sup>1,†</sup>, David Sánchez-Rodríguez <sup>1,2,\*,†</sup>, Itziar Alonso-González <sup>1,2,†</sup> and Jesús B. Alonso-Hernández <sup>2,3,†</sup>

<sup>1</sup> Telematic Engineering Department, University of Las Palmas de Gran Canaria, Campus Universitario de Tafira, 35017 Las Palmas de Gran Canaria, Spain; mangel.quintana@ulpgc.es (M.A.Q.-S.); itziar.alonso@ulpgc.es (I.A.-G.)

<sup>2</sup> Institute for Technological Development and Innovation in Communications, University of Las Palmas de Gran Canaria, Campus Universitario de Tafira, 35017 Las Palmas de Gran Canaria, Spain

<sup>3</sup> Signal and Communications Department, University of Las Palmas de Gran Canaria, Campus Universitario de Tafira, 35017 Las Palmas de Gran Canaria, Spain; jesus.alonso@ulpgc.es

\* Correspondence: david.sanchez@ulpgc.es; Tel.: +34-928-458047; Fax: +34-928-451380

† These authors contributed equally to this work.

Received: 31 July 2017; Accepted: 25 August 2017; Published: 27 August 2017

**Abstract:** Ambient Assisted Living (AAL) has become an attractive research topic due to growing interest in remote monitoring of older people. Development in sensor technologies and advances in wireless communications allows to remotely offer smart assistance and monitor those people at their own home, increasing their quality of life. In this context, Wireless Acoustic Sensor Networks (WASN) provide a suitable way for implementing AAL systems which can be used to infer hazardous situations via environmental sounds identification. Nevertheless, satisfying sensor solutions have not been found with the considerations of both low cost and high performance. In this paper, we report the design and implementation of a wireless acoustic sensor to be located at the edge of a WASN for recording and processing environmental sounds which can be applied to AAL systems for personal healthcare because it has the following significant advantages: low cost, small size, audio sampling and computation capabilities for audio processing. The proposed wireless acoustic sensor is able to record audio samples at least to 10 kHz sampling frequency and 12-bit resolution. Also, it is capable of doing audio signal processing without compromising the sample rate and the energy consumption by using a new microcontroller released at the last quarter of 2016. The proposed low cost wireless acoustic sensor has been verified using four randomness tests for doing statistical analysis and a classification system of the recorded sounds based on audio fingerprints.

**Keywords:** wireless acoustic sensor; ambient assisted living; internet of things; edge computing; low cost; ESP32

## 1. Introduction

As one of the fastest growing technologies in the emerging Internet of Things (IoT) environment, low power wireless sensor networks are expected to realize IoT applications and to provide connectivity for remote smart objects. The basic concept of IoT is that various smart objects can be automatically linked into a network for interacting with humans through perception and networking technologies [1]. Smart objects in the IoT have the ability to send information through the Internet to provide the interaction among multiple things and people. IoT is opening tremendous opportunities for novel applications that promise to improve the quality of people life.

The development of IoT technologies can be applied to a huge variety of applications, such as intelligent power grid [2], healthcare [3], environmental monitoring [4], localization [5], etc. In an

AAL context where assisted living technologies are based on ambient intelligence, smart objects need to use wireless communications because of the requirements of supporting mobile applications and remote monitoring of people. AAL can be used for detecting and preventing distress situations, improving wellness and health conditions of older adults. AAL technologies can also provide more safety for the elderly, using mobile emergency response systems, detecting domestic accidents, monitoring activities of daily living, issuing reminders, as well as helping with mobility and automation, and, overall, improving their quality of life [6,7]. In fact, according [8], AAL should be understood as a system for extending the time people can live in their preferred environment by increasing their autonomy, self-confidence and mobility; supporting the preservation of health and functional capabilities of the elderly, promoting a better and healthier lifestyle for individuals at risk; enhancing security, preventing social isolation and supporting the preservation of the multifunctional network around the individual; supporting carers, families and care organizations; and increasing the efficiency and productivity of used resources in the ageing societies.

A survey of sensors in assisted living of older people is presented in [9], such as passive infrared (PIR) and vibration sensors, accelerometers, cameras, depth sensors, and microphones. These systems should satisfy some requirements as: low-cost, high accuracy, user acceptance and privacy. These can be connected to form a network for an intelligent home designed for elderly people. The data and decision results that the sensors produce can be processed and fused over a cloud or a fog. Authors expect that the IoT will lead to remote health monitoring and emergency notification AAL systems that will operate autonomously, without requiring user intervention. In this context, audio recognition is also a promising way to ensure more safety by contributing to detection of distress situations because of the interaction of each person with her environment may be identified. In fact, in [10] detection of distress situations and monitoring of activity and health are described as two challenges to address in AAL environments. On the one hand, the identification the sounds of everyday life can be particularly useful for detecting distress situations in which the person might be. For instance, the detection of a fall can be used to call an emergency number. On the other hand, audio processing can be quite useful for the monitoring of the person's activity and the assessment of some decline. For instance, an application might consist of recognising health related symptoms such as coughing, scraping throats and sniffles. Hence, the development of WASN with low power consumption and low cost are suitable for implementing AAL systems. In this research, we are focused in the development of a low cost wireless acoustic sensor with audio processing capabilities and network connectivity to be located at the edge of a WASN.

The WASN have been developed under the paradigms of both the Smart City and the IoT. In recent years, there has been a rapid evolution of WASN, and many works have been developed. To date, several authors have designed and deployed WASN for different purposes such as noise monitoring [11] or sound identification as road traffic, horns, and people [12]. For instance, in [13] the production and analysis of a real-life environmental audio database in two urban and suburban scenarios corresponding to the pilot areas of the DYNAMAP project was presented. The WASN of the DYNAMAP project is based on low cost acoustic sensor but it is connected to digital recorder for data saving. Hence, unlike our research, audio samples cannot be sent to a central server using wireless communications, and neither audio processing can not be carried out at node. Audio recordings have been categorized as road traffic noise, background city noise, and anomalous noise events. However, it was carried out offline with Audacity and Matlab software [14].

In [15], a distributed noise measurement system based on IoT technology was developed. The sensor node is based on a Raspberry Pi with an electret omnidirectional microphone and a sound card in order to record the audio. The data from WASN was interpolated for obtaining a spatial noise level in a small-sized city. However, the system was designed to measure, represent and predict urban noise levels, and not for audio processing and classification.

In [16], the design of low cost wireless sensor node for environmental noise measurement is described. The sensor node platform is built on ATmega128L with 4 kB RAM, and its internal

10-bit ADC can operate a peak sampling rate of 33 kHz. However, according the microcontroller specifications, the maximum sampling rate for 10-bit resolution is 9.8 kHz and not 33 kHz. In addition, only the effective sound pressure is sent, and an audio processing is not carried out.

Nevertheless, the WASN paradigm presents several challenges, ranging from those derived from the design and development of the wireless sensor network, such as energy harvesting and low cost hardware development and maintenance, to some specific challenges derived from the automation of the data collection and subsequent signal processing, such as to detect anomalous noise events [13]. In addition, the sensor of a WASN designed for AAL systems should process the environmental sounds to rapidly infer hazardous situations instead to send the full audio record to a server for a centralized processing. Thus, processing data at the node can ensure shorter response time and better reliability. In this context, the use of devices with an increasing storage and computation capacity coins a new term: Edge or Fog Computing. This model extends Cloud computing and services to the edge of the network reducing network latency and offloading computation [17], as well as to avoid bottlenecks at remote server due to the throughput and volume of data to be collected and processed. Edge Computing has the potential to address the concerns of response time requirement, battery life constraint, bandwidth cost saving, as well as data safety and privacy [18]. This concept covers computational to be performed at the edge of the network and to exchange data from or to cloud IoT services. In [19], the design and deployment of a WASN at home, inspired by the Mobile Edge Computing paradigm [20] able to gather the data of all acoustic sensing nodes deployed to infer the audio events of interest in an AAL context is described. It follows a distributed implementation, where all digital signal processing is carried out in a concentrator offloading the sensor nodes and avoiding the use of the server to remotely process the data. This concentrator is based on a GPU embedded platform.

As has been discussed, many research using low cost sensors in a WASN have been developed. Nevertheless, those works have been designed to measure only noise levels and not for sound identification. On the other hand, research where audio processing is carried out are based on medium cost platforms, such as Raspberry or GPU, or using cloud services. Hence, the aim of this research is to solve these deficiencies designing a low cost acoustic sensor to do audio processing at the edge of network.

There is no doubt that significant progress has been made in the field of wireless acoustic sensor networks. However, an improvement to the actual sensors is needed because the main drawback of these recent WASN is that their power consumption and cost do not fit some of the critical requirements of AAL applications: power consumption, mobility, size and cost. In addition, humans are most sensitive to frequencies between 2 kHz and 5 kHz, and the speech and environmental sounds are often less than 5 kHz bandwidth. Therefore, the sensor has to be characterized by a spectrum with these frequencies. In this paper, a novel wireless acoustic sensor is proposed. The main novelty of this work comes from the fact that the proposed wireless acoustic sensor is able to record audio samples at least at 10 kHz sampling frequency (5 kHz bandwidth) and 12-bit resolution, and audio signal processing can be carried out at node without compromising the sample rate and the energy consumption. Furthermore, this sensor can be applied in AAL systems for personal healthcare because it has the following significant advantages: low cost, small size, wireless network connectivity, audio sampling and computation capabilities for audio processing. Thus, the identification of sounds for an AAL context, such as fall detection or health related symptoms, could be carried out at the edge of a WASN reducing network latency and improving response time and battery life of proposed sensor, enhancing quality of life and safety of older people.

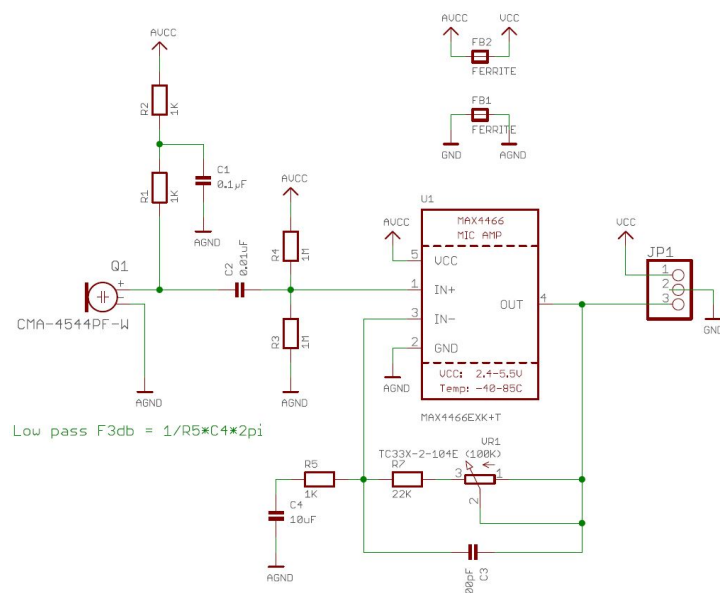
The remainder of this paper is organized as follows. Section 2 describes the low cost proposed wireless acoustic sensor. Section 3 describes the used methods to validate and evaluate the proposed sensor. Section 4 shows the results of some experiments carried out to validate the designed sensor in this study. Finally, in Section 5 draws some conclusions and discusses some possible directions for future research.

## 2. Wireless Acoustic Sensor

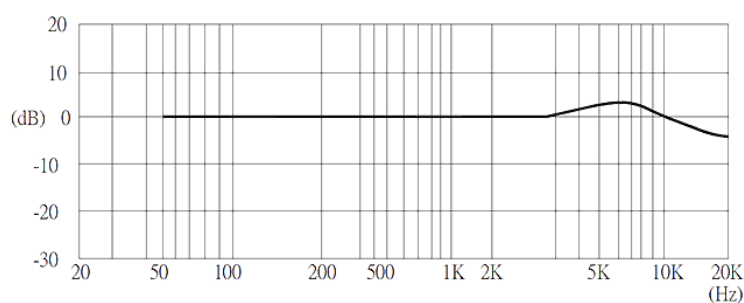
In this section the proposed low cost wireless acoustic sensor is described which is formed by an audio sensor and a microcontroller based board. The main goal in designing the sensor was to obtain a product of small size, low cost, low consumption and versatile which allows to be used in permanent and remote monitoring in AAL systems.

### 2.1. Audio Sensor

The audio sensor is an electret microphone amplifier with adjustable gain [21]. It is based on an electret omnidirectional microphone, CMA-4544PF-W, and an op-amp specifically optimized for use as microphone preamplifiers, a Maxim MAX4466, Figure 1. They provide the ideal combination of an optimized gain bandwidth product with low voltage operation in ultra-small packages. Furthermore, it has an almost flat response in the frequency range between 50 Hz and 10 kHz, Figure 2. Therefore, the characterization of sensor is fulfilled whose operating frequency lies within the range 100 Hz–5 kHz.



**Figure 1.** Schematic of audio sensor.

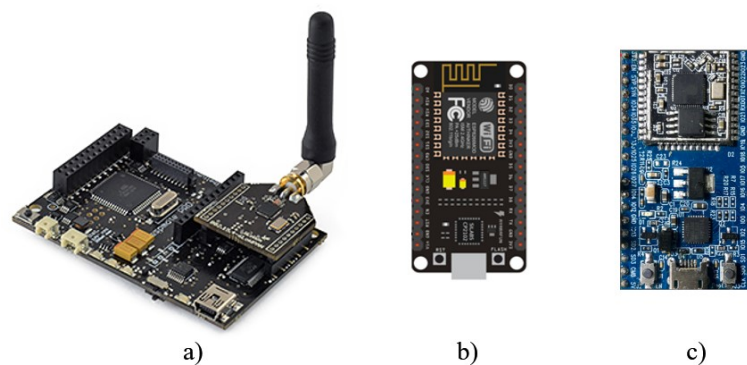


**Figure 2.** Frequency response curve of microphone.

### 2.2. Microcontroller Based Board

Three low cost microcontroller platforms were evaluated, jointly with the above audio sensor, to determine the best option for the proposed system: Libelium Wasp mote platform [22],

Espressif ESP8266 board [23], and Espressif ESP32 board [24]. Figure 3 shows the evaluated microcontroller boards.



**Figure 3.** Microcontroller boards: (a) Waspote; (b) ESP8266; (c) ESP32.

Waspote board is a modular device that allows us to install different sensors and different radio transceivers. Waspote hardware architecture has been specially designed to be extremely low consumption. The Waspote has an Atmega1281 microcontroller running at 14 MHz with programmable sleep modes. These sleep modes make Waspote the lowest consumption sensor platform in the market (0.7  $\mu$ A in hibernate mode and 55  $\mu$ A in sleep mode). The whole set, formed by Waspote and audio sensor, has a small size (85  $\times$  75  $\times$  35 mm, included battery). The ATmega1281 has 8 ADC channels with 10-bit resolution. Due to the microcontroller characteristics, the tested maximum ADC sampling frequency was 9.8 kHz. In addition, it has only 8 kB SRAM, and therefore, the audio recording is about a few tenths of a second maximum duration. Waspote has an SD card and could be used to save the sampled data. Nevertheless, the sample rate of ADC converter must be fit to 8-bit resolution to carry out these extra operations needed.

ESP8266 board delivers a highly integrated Wi-Fi SoC solution for efficient power, with its complete and self-contained Wi-Fi networking capabilities. It integrates an enhanced version of Tensilica's L106 Diamond series 32-bit processor and on-chip SRAM with an ADC with 10-bit resolution, and can be interfaced with external sensors through the GPIOs, in low development cost at prototyped. One of the most common boards with the ESP8266 is NodeMCU, with ESP-12E module, Figure 3b. The whole set, formed by ESP8266 and audio sensor, has a very small size (50  $\times$  30  $\times$  20 mm) which is very useful to place at different positions in a discrete way. It can support up to 80 MHz frequency clock. It has a built-in SPI flash memory with 4MB capacity and the SRAM capacity available to users is about 36 kB. The tested maximum ADC sampling frequency was 10.6 kHz.

Espressif Systems announced the launch of ESP32 cloud on chip on September 6th, 2016. It is a Dual Core Wi-Fi + BT Combo MCU. Some of features of the ESP32 are the following: the CPU is an Xtensa Dual-Core 32-bit LX6 microprocessor, operating up to 240 MHz, 520 kB SRAM, 12-bit SAR ADC up to 18 channels and built-in Wi-Fi card, supporting IEEE 802.11 b/g/n standards, and Bluetooth v4.2 BR/EDR and BLE. Also, the ESP32 chip features 40 physical GPIO pads which can be used as a general purpose I/O, to connect new sensors, or can be connected to an internal peripheral signal. The most common development board is the ESP32S, with a ESP-WROOM-32 module and an SRAM capacity available to users about 300 kB. As previous board, the whole set, formed by ESP32 and audio sensor, has a very small size (55  $\times$  30  $\times$  20 mm), and is very useful to place at different positions in a discrete way. The tested maximum ADC sampling frequency was 100 kHz. It is enough for the system purposes. Furthermore, ADC is non-blocking, so the conversion process with other instructions execution can be overlapped.

Although all analyzed boards can fulfill for implementing a wireless acoustic sensor, ESP32 board was chosen because it has the biggest: ADC bit resolution, SRAM capacity, and microprocessor frequency. Furthermore, because of it has a dual core and the microcontroller's connectivity features



and functionalities, audio samples are gathered while other operations can be simultaneously done, such as sending data to a server using IP protocol or data processing, thus promoting the edge computing idea. Figure 4 shows the wireless acoustic sensor based on ESP32 board. On the other hand, the cost of proposed sensor is about 10 Euros, being it very competitive for an AAL environment.

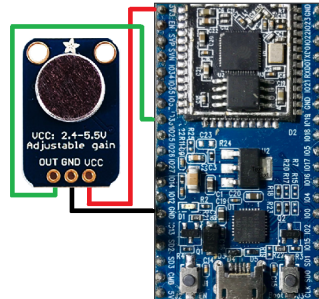


Figure 4. Wireless acoustic sensor based on ESP32 board.

Lastly, the principle of operation of the software implementation for the proposed wireless acoustic sensor is shown in Figure 5. First, a timer interruption is enabled for gathering the audio samples from ADC using *AnalogRead* function. Timer is set to 100  $\mu$ s to obtain a sampling frequency at 10 kHz. Next, the data obtained from ADC are stored in an endless buffer. Finally, the raw data can be sent to a server for recording the sampled data in a wav format file or a suitable audio fingerprint to identify different sound events that can be used for detecting hazardous situations.

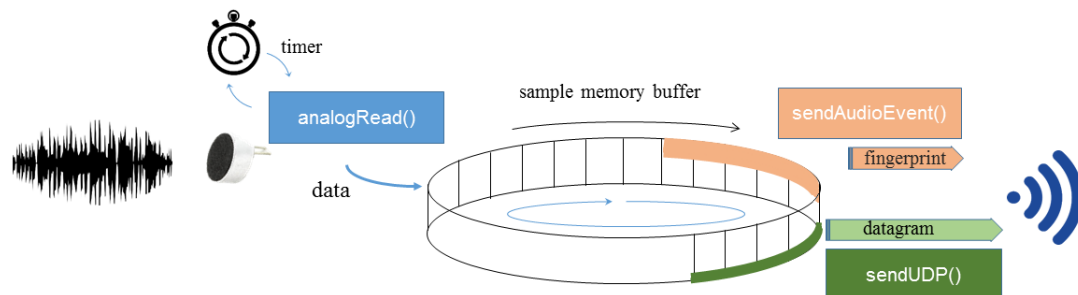


Figure 5. Graphical flow diagram implemented in the wireless acoustic sensor.

### 3. System Validation Methods

In order to validate the proposed wireless acoustic sensor a statistical analysis and an audio classification of recorded samples are carrying out. Thus, randomness tests and an audio fingerprint matching were the methods employed for system validation and are described in this section.

#### 3.1. Randomness Tests

Randomness tests can be used to determine whether a dataset has a recognizable pattern, and therefore whether the process that generated it is significantly random. That is, it can be used to test the hypothesis that the elements of a sequence are mutually independent or not. Four randomness tests were used to demonstrate that recorded audio files by the proposed system have a recognizable pattern, and hence, the sampled audio information is not random. The following randomness tests were used: Bartels Test [25], Cox Stuart Test [26], Mann-Kendall Test [27] and Wald-Wolfowitz Test [28].

##### 3.1.1. Bartels Test

This randomness test is the rank version of von Neumann's Ratio Test for Randomness [29].

### 3.1.2. Cox Stuart Test

In this test data are grouped in pairs with the  $i$ th observation of the first half paired with the  $i$ th observation of the second half of the time-ordered data. If the length of vector  $X$  is odd the middle observation is eliminated. The Cox Stuart test is then simply a sign test applied to these paired data.

### 3.1.3. Mann-Kendall Test

This randomness test is a non-parametric statistical test that analyzes difference in signs between earlier and later data points. The idea is that if a trend is present, the sign values will tend to increase constantly, or decrease constantly. Every value is compared to every value preceding it in the time series.

### 3.1.4. Wald-Wolfowitz Test

This randomness test is a non-parametric statistical test that transforms into a dichotomous vector according as each values is above or below a given threshold. Values equal to the level are removed from the sample. The default threshold value used in applications is the sample median.

## 3.2. Audio Fingerprint Matching

An audio fingerprint is a compact content-based signature that summarizes an audio recording. This technology has attracted attention since they allow the identification of audio independently of its format and without the need of meta-data or watermark embedding [30]. The main objective of an audio fingerprint mechanism is to efficiently compare the equality (or not) of two audio files, not by comparing the files themselves, but by comparing substantially smaller sets of information, referred to as audio fingerprints. Furthermore, audio fingerprint length is a lot less than the raw audio data. In order to validate the proposed wireless acoustic sensor, an open source application, termed Chromaprint [31], is used to generate the fingerprints of original and recorded audios. Then, to find an audio matching between original and recorded audios, the Hamming distance is evaluated using both fingerprints.

### 3.2.1. Chromaprint Process

Chromaprint converts the audio input to mono and downsampled to 11,025 Hz. The audio signal is converted to the frequency domain by performing a short-time Fourier Transform (STFT) with a frame size of 4096 samples (371 ms) and a 2/3 overlap (2731 samples). The resulting spectrum is converted to 12 bins representing the chroma of the signal. This information is called “chroma features”. Each bin in the chromagram represents the energy that is present in a musical note. The 12 bins represent the 12 notes of the chromatic scale. In order to transform the bins in a more compact form to carry out the fingerprint matching, a 12-by-16 sliding window is moved over the chromagram one sample at a time. On each of them is applied a pre-defined set of 16 filters that capture intensity differences across musical notes and time. Each of the filters quantizes the energy value to a 2-bit number. The 2-bit value is encoded using Gray coding. The 2-bit hash values from each of the 16 filters are converted to a single 32-bit integer representing the subfingerprint of the 12-by-16 window. The window is advanced one sample to calculate the next subfingerprint. The full fingerprint is composed by the all subfinngerprints.

### 3.2.2. Hamming Distance

In order to find a simple audio matching for verifying and validating the proposed system, the Hamming distance is implemented because is performed at the bit-level and therefore, requires less computational complexity. The Hamming distance between two  $(NF \times 32)$ -bit binary fingerprint vectors  $fv_1$  and  $fv_2$  is computed as Equation (1):

$$H_d(fv_1, fv_2) = \sum_{i=1}^{NF \times 32} F(bit_{fv_1}(i) \neq bit_{fv_2}(i)) \quad (1)$$

where  $NF$  denotes the number of subfingerprints of vectors,  $bit_{fv_1}(i)$  and  $bit_{fv_2}$  are the  $i$ th element of binary fingerprint vectors, and  $F()$  is an function defined by Equation (2):

$$F(x) = \begin{cases} 1 & \text{if } x \text{ is true} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

### 3.2.3. Matching Algorithm

Algorithm 1 was used to evaluate the shortest Hamming distance between all original environmental sounds and each recorded audio. The shortest distance identifies and matches the recorded audio with the original sound.

---

#### Algorithm 1 Audio Fingerprint Matching

---

**Require:** Fingerprint of recorded audio,  $fv_1$

**Require:** Fingerprint of all original audios,  $fv_{source}$

$L1 \leftarrow \text{length}(fv_1)$

**for**  $source \leftarrow 1$ : all original fingerprints **do**

$L_{source} \leftarrow \text{length}(fv_{source})$

**for**  $i \leftarrow 1 : (L_{source} - L1 - 1)$  **do**

$distanceVector \leftarrow H_d(fv_1, fv_{source}(i : L1 + i - 1))$

**end for**

$distanceAllAudioVector \leftarrow \min(distanceVector)$

**end for**

**return**  $\min(distanceAllAudioVector)$

---

## 4. Results and Discussion

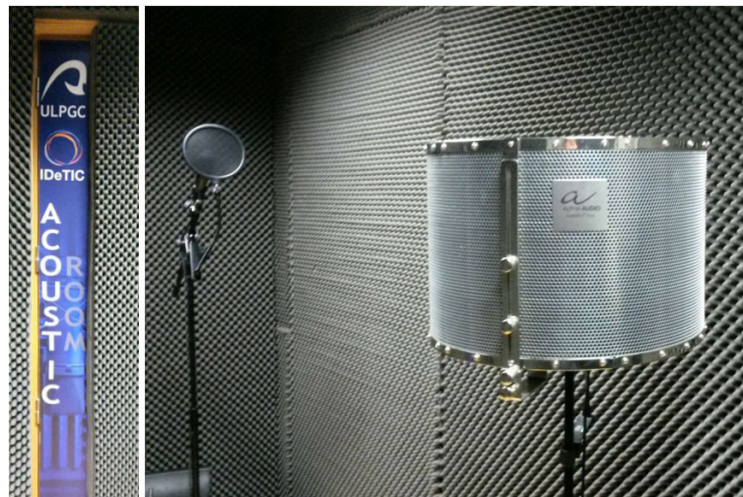
This section describes the acoustic anechoic chamber where different audio samples from different environmental sounds were gathered and the dataset built to evaluate the validity of the proposed sensor. Afterwards, the results of aforementioned system evaluation methods, randomness tests and audio fingerprint matching, are presented and discussed. Also, the usefulness of the proposed sensor board in terms of energy consumption and audio processing capabilities are carried out.

### 4.1. Acoustic Anechoic Chamber

The acoustic anechoic chamber where experiments were carried out is located on the second floor of Institute for Technological Development and Innovation in Communications (IDeTIC) building at Las Palmas de Gran Canaria University, Spain. The chamber area is nearly 200 cm wide and 430 cm long, and it has a simple design to absorb reflections of sound waves and is also isolated from waves entering from its surroundings.

The soundproofing of the chamber is carried out using foam pyramidal panels which is a powerful sound absorber that dramatically reduces echo, reverberation and standing waves. For the acoustic insulation is used rock wool and polyurethane panels. Figure 6 shows acoustic anechoic chamber at IDeTIC and the foam pyramidal panels used.





**Figure 6.** Acoustic anechoic chamber at IDeTIC.

#### 4.2. Dataset

A total of 48 audio records were gathered from fourteen different indoor and outdoor environmental sounds for performing the statistical analysis and audio classification. These environmental records have been downloaded from Freesound website [32]. Table 1 shows the dataset characteristics. Each environmental sound was recorded using a 10 kHz sampling frequency and 8-bit resolution during 10 s. In order to gather different samples, the start point of each recording was randomly established.

**Table 1.** Audio recordings dataset.

Environmental Sound	Duration (s)	Number of Records
S1—Traffic jam in a city	49	3
S2—People on a street without traffic	34	3
S3—Very strong traffic	70	5
S4—City park with children	33	3
S5—Pedestrian zone of a city with traffic	32	3
S6—Inside of a noisy room by traffic	60	4
S7—Ambulance passing with the siren	29	4
S8—Drilling machine in a city	18	3
S9—Police car passing with the siren	28	3
S10—Ambulance siren. Doppler effect	24	3
S11—Dense traffic in a city	72	5
S12—Indoor door slam	23	3
S13—Indoor gun shots	98	3
S14—Slicing vegetables in a kitchen	40	3

##### 4.2.1. Randomness Tests

All recorded audios were evaluated with the four above randomness tests. The null hypothesis of randomness is tested against nonrandomness, and a  $p$ -value is calculated which is used in the context of null hypothesis testing in order to quantify the idea of statistical significance of evidence, that is, the probability of finding the observed results when the null hypothesis is true. In the tests, if the  $p$ -value is less than 0.05, the null hypothesis is rejected because a significant difference exists.

Table 2 shows the  $p$ -values of four randomness tests. As can be seen, all  $p$ -values for Bartels, Mann-Kendall and Wald-Wofowitz tests are less than 0.05. For Cox Stuart test, most results return a  $p$ -value less than 0.05, and only eight tests are slightly greater than this value. It is not significant,

and hence the null hypothesis can be rejected. Thus, it can be considered that the recorded audio files by wireless acoustic sensor are not significantly random, and therefore, have a recognizable pattern.

**Table 2.** *p*-values of randomness tests.

Record	Bartels	Cox Stuart	Mann-Kendall	Wald-Wolfowitz
S1-R1	0	0.06	≈0	0
S1-R2	0	0.01	≈0	0
S1-R3	0	0.09	≈0	0
S2-R1	0	≈0	≈0	0
S2-R2	0	≈0	≈0	0
S2-R3	0	≈0	≈0	0
S3-R1	0	≈0	≈0	0
S3-R2	0	0.02	≈0	0
S3-R3	0	0.12	≈0	0
S3-R4	0	0.13	≈0	0
S3-R5	0	0.09	≈0	0
S4-R1	0	≈0	0	0
S4-R2	0	≈0	0	0
S4-R3	0	≈0	≈0	0
S5-R1	0	≈0	≈0	0
S5-R2	0	≈0	0	0
S5-R3	0	≈0	0	0
S6-R1	0	≈0	0	0
S6-R2	0	≈0	0	0
S6-R3	0	≈0	0	0
S6-R4	0	≈0	0	0
S7-R1	0	0.07	0	0
S7-R2	0	0.08	≈0	0
S7-R3	0	0.02	≈0	0
S7-R4	0	≈0	≈0	0
S8-R1	0	≈0	≈0	0
S8-R2	0	0.01	≈0	0
S8-R3	0	≈0	≈0	0
S9-R1	0	≈0	0	0
S9-R2	0	≈0	0	0
S9-R3	0	0.01	0	0
S10-R1	0	0.03	≈0	0
S10-R2	0	0.03	≈0	0
S10-R3	0	0.06	≈0	0
S11-R1	0	≈0	≈0	0
S11-R2	0	0.01	≈0	0
S11-R3	0	≈0	≈0	0
S11-R4	0	0.06	≈0	0
S11-R5	0	≈0	≈0	0
S12-R1	0	0.01	≈0	0
S12-R2	0	0.01	0	0
S12-R3	0	≈0	0	0
S13-R1	0	≈0	0	0
S13-R2	0	≈0	0	0
S13-R3	0	0.01	0	0
S14-R1	0	0.05	0	0
S14-R2	0	≈0	0	0
S14-R3	0	0.06	0	0

#### 4.2.2. Audio Fingerprint Matching

To carry out the audio fingerprint matching, an audio fingerprint was computed both for each recorded audio and original environmental sounds. All recorded audios are 10 s long, therefore, each recorded audio fingerprint is 66 subfingerprints long. Thus, the vectors used in Equation (1)

are 2112 ( $66 \times 32$ ) bits length. Afterwards, Algorithm 1 was used to evaluate the shortest Hamming distance between all original environmental sounds and each recorded audio. The shortest distance identifies and matches the recorded audio with the original sound.

Table 3 shows the results returned by Algorithm 1 when it was evaluated for each recorded audio. For each one the shortest Hamming Distance is bold marked. As can be seen, most of recorded audios match with its correspondent original sound, yielding an 85.4% accuracy, and there is not dependency with the kind of sound, indoor or outdoor. Furthermore, a 91.6% accuracy is reached when the three shortest Hamming distances are used, that is, a recorded audio is correctly classified and found in a set of three sounds with a 91.6% probability. On the other hand, all S10 recordings were classified as S3 original audio. It can be because doppler effect is perceived in both audios. In any case, the aim of this experiment is not to implement a robust classification system, but to demonstrate the validity of the proposed acoustic sensor. Taking in account the results, it can be asserted that the recorded audios have a high grade of similarity with its original sound, and hence, the proposed acoustic sensor can be validated.

Table 3. Audio fingerprints matching.

Record	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14
S1-R1	<b>773</b>	1088	1075	1086	1132	1107	1249	1232	1096	1257	1111	1232	1060	1146
S1-R2	<b>1025</b>	1086	1047	1050	1111	1182	1323	1121	1146	1164	1053	1234	1040	1239
S1-R3	<b>883</b>	1012	1022	996	1151	1108	1227	1152	1111	1180	1101	1252	923	1114
S2-R1	1058	<b>351</b>	995	1017	1087	1100	1237	1118	1138	1208	997	1170	855	1156
S2-R2	1047	<b>336</b>	961	962	1083	1120	1231	1150	1161	1245	1022	1091	838	1127
S2-R3	1093	1088	<b>1083</b>	<b>1070</b>	1124	1224	1265	1162	1230	1288	1105	1204	1098	1194
S3-R1	1115	992	<b>549</b>	1002	1111	1037	1288	1212	1099	1298	1072	1150	890	1034
S3-R2	1144	995	<b>486</b>	1053	1121	1110	1253	1277	1127	1289	1102	1138	924	1116
S3-R3	1138	1037	<b>459</b>	1053	1223	1126	1225	1247	1178	1290	1097	1203	927	1144
S3-R4	1122	1036	<b>427</b>	1035	1117	1091	1303	1116	1179	1276	1081	1081	962	1060
S3-R5	1112	1066	<b>485</b>	1126	1180	1177	1247	1198	1168	1276	1061	1177	1023	1183
S4-R1	1032	939	1003	<b>370</b>	1133	1051	1250	1157	1160	1249	1064	1195	833	1159
S4-R2	1035	978	996	<b>396</b>	1107	1133	1187	1159	1135	1270	1061	1150	867	1142
S4-R3	1102	1051	1089	<b>313</b>	1122	1128	1280	1182	1218	1172	1116	1211	951	1164
S5-R1	1112	1112	1060	1134	<b>377</b>	1102	1215	1157	1178	1262	1088	1162	1081	1163
S5-R2	1095	1024	1033	1132	1053	1064	1246	1170	1119	1376	1058	1143	<b>973</b>	1121
S5-R3	1140	1080	1108	1182	<b>403</b>	1108	1236	1174	1204	1270	1135	1193	1102	1150
S6-R1	1209	1171	1153	1161	1131	<b>591</b>	1240	1198	1318	1306	1155	1188	1172	1010
S6-R2	1141	1019	1080	1141	1059	<b>598</b>	1248	1149	1191	1277	1099	1166	1022	1066
S6-R3	1043	1035	<b>972</b>	1051	1107	1040	1365	1106	1088	1152	1073	1153	1055	1059
S6-R4	1134	1051	1026	1164	1093	<b>485</b>	1289	1151	1172	1275	1152	1185	1007	1039
S7-R1	1167	1121	1089	1153	1150	1178	<b>1026</b>	1220	1196	1170	1145	1095	1059	1129
S7-R2	1208	1217	1091	1215	1247	1205	<b>991</b>	1226	1209	1234	1183	1179	1183	1169
S7-R3	1236	1192	1133	1256	1273	1225	<b>340</b>	1371	1283	1311	1166	1159	1335	1212
S7-R4	1201	1174	1136	1165	1216	1174	<b>351</b>	1246	1247	1240	1151	1196	1242	1227
S8-R1	981	1096	1155	1069	1096	1196	1265	<b>949</b>	1091	1079	977	1186	982	1218
S8-R2	1015	1111	1123	1047	1036	1161	1274	<b>340</b>	1095	1077	986	1181	986	1236
S8-R3	1007	1093	1136	1071	1081	1189	1289	<b>847</b>	1079	1027	975	1217	988	1238
S9-R1	1113	1138	1111	1133	1151	1036	1380	1056	<b>1040</b>	1103	1099	1167	1098	1018
S9-R2	1164	1195	1143	1245	1192	1061	1322	1156	<b>402</b>	1228	1128	1199	1106	1052
S9-R3	1140	1129	1142	1191	1199	1160	1314	1139	<b>445</b>	1195	1110	1231	1052	1193
S10-R1	1151	1118	<b>1065</b>	1160	1275	1201	1227	1355	1186	1380	1160	1197	1194	1097
S10-R2	1149	1061	<b>1060</b>	1124	1173	1105	1322	1250	1073	1305	1157	1219	1192	1128
S10-R3	1191	1140	<b>1092</b>	1095	1205	1200	1206	1218	1152	1299	1112	1237	1175	1148
S11-R1	1072	1031	1105	1071	1106	1147	1259	1121	1136	1334	<b>483</b>	1141	903	1125
S11-R2	1070	1084	1085	1059	1125	1112	1207	1090	1129	1287	<b>538</b>	1238	981	1162
S11-R3	997	1066	1121	1051	1080	1138	1280	1030	1111	1190	<b>483</b>	1175	922	1249
S11-R4	1144	1174	1161	1208	1132	1095	1187	1192	1179	1311	<b>358</b>	1167	1136	1156
S11-R5	1066	1022	1082	1038	1125	1086	1261	1163	1178	1340	<b>488</b>	1170	931	1068
S12-R1	1227	1222	1195	1219	1166	1140	1181	1263	1265	1297	1194	<b>495</b>	1288	1192
S12-R2	1094	1055	1051	1033	1125	1115	1252	1029	1102	1193	1080	1024	<b>953</b>	1183
S12-R3	1121	1095	1122	1094	1140	1156	1262	1122	1233	1228	1054	<b>1009</b>	1074	1149

Table 3. Cont.

Record	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14
S13-R1	1039	1031	991	1019	1110	1076	1400	972	1108	1031	985	1169	<b>758</b>	1060
S13-R2	1029	966	956	942	1140	1036	1323	1036	1015	1128	986	1165	<b>618</b>	1058
S13-R3	953	954	994	860	1099	1072	1246	1045	1055	1142	937	1217	<b>372</b>	1180
S14-R1	1082	1052	1077	1111	1113	1148	1293	1071	1102	1140	1120	1158	985	<b>758</b>
S14-R2	1078	979	1022	1085	1106	1095	1295	1093	1122	1172	1117	1159	1001	<b>768</b>
S14-R3	1244	1158	1129	1228	1170	931	1381	1233	1110	1360	1219	1092	1156	<b>249</b>

#### 4.3. Energy Consumption and Audio Processing Capabilities

In order to evaluate the energy consumption of the proposed sensor, three experiments with different processing capabilities were performed: (A) audio recording, (B) audio recording and Fast Fourier Transform (FFT) calculation, and (C) audio recording and UDP datagram sending via Wi-Fi connection. Audio recording was carried out in an infinite loop using a 10 kHz sampling frequency and 12-bit resolution. The ArduinoFFT library [33] was used to implement the FFT, and it was computed each 12.8 ms, that is, every FFT was run after 128 new samples were recorded. UDP datagram sending was carried out each 25.6 ms, therefore, each UDP datagram is sent when 256 new samples are gathered. The prototype sensor was powered by 5 V, and the current consumption was measurement for each experiment. Table 4 shows the results. As can be seen, the A and B experiments have similar energy consumption because all operations are carried out in an infinite loop and additional resources are not used, only the processor. However, in the C experiment, Wi-Fi module is periodically transmitting a datagram, and therefore, the energy consumption is higher. In any case, the maximum energy consumption is about 0.8 W, and the proposed sensor could be powered by battery for a long time.

Table 4. Energy consumption.

Experiment	Average Current (mA)	Energy Consumption (W)
A—Audio recording	139	0.695
B—Audio recording and FFT	141	0.705
C—Audio recording and UDP sending	165	0.825

On the other hand, audio processing capabilities were evaluated implementing the FFT with different number of samples and carrying out an audio recording using the same core. FFT was chosen because is an expensive computational algorithm in audio processing. Each experiment was performed 1000 times and the average execution time was computed. Table 5 shows the average execution time for each experiment. As can be seen, the execution time increases with the number of samples. In addition, using the same number of samples, the FFT execution time is slightly higher when the simultaneous sampling is carried out, that is, the FFT was computed while other samples were gathered. Nevertheless, about 20 ms is only spent to compute a 512 sample FFT. Hence, audio processing capabilities could be performed without compromising the sample rate and the energy consumption, and edge computing paradigm can be implemented in the proposed sensor.

Table 5. Fast Fourier Transform execution time.

Number of Samples	Simultaneous Sampling	Average Execution Time (ms)
128	No	3.32
128	Yes	4.45
256	No	7.04
256	Yes	9.43
512	No	14.92
512	Yes	19.97

## 5. Conclusions and Future Work

As was discussed in this paper, in recent years, many authors have experienced a growing interest in remote monitoring of older people, and several systems have been proposed in the literature. These systems are commonly termed as AAL systems, but the acoustic sensors were not designed with low cost and audio processing requisites. In this paper, we described the design of a low cost wireless acoustic sensor for AAL systems based on ESP32 board. In order to choose the best platform, three different low cost microcontroller boards were evaluated. It was given a detailed description of the hardware and the principle of operation of software implementation. The proposed sensor is capable of recording ambient sounds at least to 10 kHz sampling frequency and 12-bit resolution. Furthermore, the sensor board has computation capabilities to carry out audio signal processing and network communications without compromising the sample rate and the energy consumption. Hence, the proposed sensor can improve AAL solutions carrying out the audio identification for monitoring of activity and health, and the detection of distress situations at the edge of WASN. Thus, a shorter response time and better reliability is ensured enhancing quality of life and safety of older people. The acoustic sensor is very small in size, and therefore is very useful to be used in a discrete way for personal healthcare in AAL systems, and the cost of hardware platform is very competitive. The experiments on the proposed system showed that the system worked well. System validation is demonstrated by experimental results, which were statistically obtained analysing several tests of randomness and audio classification. Furthermore, evaluations of energy consumption and audio processing capabilities were carried out demonstrating the usefulness and low power, and that edge computing paradigm can be implemented at the proposed sensor.

In our ongoing work, we are planning to design a better sound classification system based on audio fingerprint to be implemented at each acoustic sensor. Moreover, we are also planning to deploy a WASN using the proposed acoustic sensor in this paper to evaluate the whole system and network performance.

**Author Contributions:** Miguel A. Quintana-Suárez designed and implemented the prototype of wireless acoustic sensor, and wrote part of manuscript. David Sánchez-Rodríguez, Itziar Alonso-González and Jesús B. Alonso-Hernández conceived and designed the experiments, analyzed the data, and wrote the rest of the manuscript. All authors have reviewed and approved the final manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Chang, C.Y.; Kuo, C.H.; Chen, J.C.; Wang, T.C. Design and implementation of an IoT access point for smart home. *Appl. Sci.* **2015**, *5*, 1882–1903.
2. Peruzzini, M.; Germani, M.; Papetti, A.; Capitanelli, A. Smart Home Information Management System for Energy-Efficient Networks. In *Collaborative Systems for Reindustrialization, Proceedings of the 14th IFIP WG 5.5 Working Conference on Virtual Enterprises, PRO-VE 2013 Dresden, Germany, 30 September–2 October 2013*; Springer: Heidelberg, Germany, 2013; pp. 393–401.
3. Amendola, S.; Lodato, R.; Manzari, S.; Occhiuzzi, C.; Marrocco, G. RFID technology for IoT-based personal healthcare in smart spaces. *IEEE Internet Things J.* **2014**, *1*, 144–152.
4. Sánchez-Rosario, F.; Sanchez-Rodriguez, D.; Alonso-Hernández, J.B.; Travieso-González, C.M.; Alonso-González, I.; Ley-Bosch, C.; Ramírez-Casañas, C.; Quintana-Suárez, M.A. A low consumption real time environmental monitoring system for smart cities based on ZigBee wireless sensor network. In *Proceedings of the 2015 International Wireless Communications and Mobile Computing Conference (IWCMC), Dubrovnik, Croatia, 24–25 August 2015*; IEEE: Washington, DC, USA, 2015; pp. 702–707.
5. Lin, K.; Chen, M.; Deng, J.; Hassan, M.M.; Fortino, G. Enhanced fingerprinting and trajectory prediction for IoT localization in smart buildings. *IEEE Trans. Autom. Sci. Eng.* **2016**, *13*, 1294–1307.
6. Memon, M.; Wagner, S.R.; Pedersen, C.F.; Beevi, F.H.A.; Hansen, F.O. Ambient assisted living healthcare frameworks, platforms, standards, and quality attributes. *Sensors* **2014**, *14*, 4312–4341.

7. Rashidi, P.; Mihailidis, A. A survey on ambient-assisted living tools for older adults. *IEEE J. Biomed. Health Inform.* **2013**, *17*, 579–590.
8. European Commission. Active and Assisted Living Programme. ICT for Ageing Well. Available online: <http://www.aal-europe.eu/> (accessed on 1 June 2017).
9. Erden, F.; Velipasalar, S.; Alkar, A.Z.; Cetin, A.E. Sensors in Assisted Living: A survey of signal and image processing methods. *IEEE Signal Proc. Mag.* **2016**, *33*, 36–44.
10. Vacher, M.; Portet, F.; Fleury, A.; Noury, N. Development of audio sensing technology for ambient assisted living: Applications and challenges. *Int. J. E-Health and Medical Commun.* **2011**, *2*, 35–34.
11. Kivelä, I.; Gao, C.; Luomala, J.; Hakala, I. Design of Noise Measurement Sensor Network: Networking and Communication Part. In Proceedings of the Fifth International Conference on Sensor Technologies and Applications, Sensorcomm, French Riviera, France, 21–27 August 2011; pp. 280–287.
12. Paulo, J.; Fazenda, P.; Oliveira, T.; Carvalho, C.; Felix, M. Framework to Monitor Sound Events in the City Supported by the FIWARE platform. In Proceedings of the 46th Congreso Español de Acústica, Valencia, Spain, 21–23 October 2015; pp. 21–23.
13. Alías, F.; Socoró, J.C. Description of anomalous noise events for reliable dynamic traffic noise mapping in real-life urban and suburban soundscapes. *Appl. Sci.* **2017**, *7*, 146.
14. Sevillano, X.; Socoró, J.C.; Alías, F.; Bellucci, P.; Peruzzi, L.; Radaelli, S.; Coppi, P.; Nencini, L.; Cerniglia, A.; Bisceglie, A.; et al. DYNAMAP—Development of low cost sensors networks for real time noise mapping. *Noise Mapp.* **2016**, *3*, 172–189.
15. Segura Garcia, J.; Pérez Solano, J.J.; Cobos Serrano, M.; Navarro Camba, E.A.; Felici Castell, S.; Soriano Asensi, A.; Montes Suay, F. Spatial Statistical Analysis of Urban Noise Data from a WASN Gathered by an IoT System: Application to a Small City. *Appl. Sci.* **2016**, *6*, 380.
16. Hakala, I.; Kivela, I.; Ihalainen, J.; Luomala, J.; Gao, C. Design of Low-Cost Noise Measurement Sensor Network: Sensor Function Design. In Proceedings of the 2010 First International Conference on Sensor Device Technologies and Applications (SENSORDEVICES), Venice, Italy, 18–25 July 2010; IEEE: Washington, DC, USA, 2010; pp. 172–179.
17. Ahmed, A.; Ahmed, E. A Survey on Mobile Edge Computing. In Proceedings of the 2016 10th International Conference on Intelligent Systems and Control (ISCO), Coimbatore, India, 7–8 January 2016; IEEE: Washington, DC, USA, 2016; pp. 1–8.
18. Shi, W.; Cao, J.; Zhang, Q.; Li, Y.; Xu, L. Edge computing: Vision and challenges. *IEEE Internet Things J.* **2016**, *3*, 637–646.
19. Alsina-Pagès, R.M.; Navarro, J.; Alías, F.; Hervás, M. homeSound: Real-Time Audio Event Detection Based on High Performance Computing for Behaviour and Surveillance Remote Monitoring. *Sensors* **2017**, *17*, 854.
20. Mach, P.; Becvar, Z. Mobile edge computing: A survey on architecture and computation offloading. *IEEE Commun. Surv. Tutor.* **2017**, *19*, 1628–1656.
21. Adafruit. Electret Microphone Amplifier. Available online: <https://www.adafruit.com/product/1063> (accessed on 15 July 2017).
22. Libelium. Waspote Platform. Available online: <http://www.libelium.com/products/waspote/> (accessed on 3 May 2017).
23. Espressif. ESP8266 Specification. Available online: <https://espressif.com/en/products/hardware/esp8266ex/overview> (accessed on 15 July 2017).
24. Espressif. ESP32 Specification. Available online: <https://espressif.com/en/products/hardware/esp32/overview> (accessed on 15 July 2017).
25. Bartels, R. The rank version of von Neumann’s ratio test for randomness. *J. Am. Stat. Assoc.* **1982**, *77*, 40–46.
26. Cox, D.R.; Stuart, A. Some quick sign tests for trend in location and dispersion. *Biometrika* **1955**, *42*, 80–95.
27. Kendall, M. *Rank Correlation Methods*; Oxford University Press: New York, NY, USA, 1990.
28. Wald, A.; Wolfowitz, J. On a test whether two samples are from the same population. *Ann. Math. Stat.* **1940**, *11*, 147–162.
29. Von Neumann, J. Distribution of the ratio of the mean square successive difference to the variance. *Ann. Math. Stat.* **1941**, *12*, 367–395.
30. Cano, P.; Battle, E.; Kalker, T.; Haitsma, J. A review of audio fingerprinting. *J. VLSI Signal Proc. Syst. Signal Image Video Technol.* **2005**, *41*, 271–284.
31. Lukáš, L. Chromaprint. Available online: <https://acoustid.org/chromaprint> (accessed on 3 May 2017).



32. Freesound. Dataset of Audio Clips. Available online: <http://freesound.org/> (accessed on 1 July 2017).
33. Condes, E. ArduinoFFT. Available online: <https://github.com/kosme/arduinoFFT> (accessed on 28 July 2017).



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).