SOUND SOURCE LOCALIZATION USING MEMS DIGITAL MICROPHONE ARRAYS

by

Mustafa Aşıroğlu

B.S., Electrical and Electronics Engineering, Boğaziçi University, 2015

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Electrical and Electronics Engineering

Boğaziçi University

2019

ACKNOWLEDGMENTS

It is a priceless present that working with thoughtful and supportive people. I had this present during this study, I would like to thank to all those brilliant people around.

I would like to start with my supervisor, Prof. Şenol Mutlu. I am very pleased to work continuously with him during for last 4 years including my undergraduate senior design project. His experience, his knowhow and constructive guidance are very valuable and inspiring for me.

I would like to thank Zahid Sağıroğlu, PhD candidate in Boğaziçi University, who is always ready to help with his brilliant mind and strong hardware experience. Another person I would like to name here is Ogün Kırmemiş for his continuous support with his strong analytical background on signal processing area.

Finally, I owe my greatest gratitude to my wife, dear Ece, for her never-ending moral support and positive energy during agonizing periods of this study. Last but not least, I would like to thank to my family, especially to my mother, for their encouragement for me to succeed this thesis work.

ABSTRACT

SOUND SOURCE LOCALIZATION USING MEMS DIGITAL MICROPHONE ARRAYS

Technological developments around artificial intelligence in the last decade have led to a significant increase in applications of audio capturing and speech processing. To increase quality of sound recording and to obtain additional spatial information from sound, microphone arrays including geometrically distributed number of microphones are used widely in last decade, but dealing with microphone arrays was requiring advanced circuitry to make digital to analog conversion and also matching circuitry to ensure synchronization and high quality. Thanks to recent advances regarding MEMS Digital Microphones, we now have very small, ultra-low power and high-quality microphone devices in the market. In this thesis, most recent MEMS digital microphone alternatives are studied to create a microphone array which is used in sound source localization and beam forming applications. Alternative digital data streaming interfaces by different manufacturers including I2S and TDM were tested and evaluated and then an evaluation PCB was produced to design a real-life MEMS digital microphone array device including 4 microphones in edges of a square shaped PCB. During evaluation part of the microphone arrays, time-delay estimation model was used as sound source localization since it is the one of the most common used models thanks to its robustness and fast performance for real-use cases including speech recognition and teleconference room systems. In order to calculate delay of each channels accurately, GCC-PHAT Cross-Correlation method was used and then direction arrival angle was obtained using array geometry inputs. Lastly, an enclosure-device box was also designed and 3D fabricated in order to eliminate possible effects of acoustic port outer design.

ÖZET

SAYISAL MEMS (MİKRO ELEKTRO-MEKANİK SİSTEMLER) MİKROFON DİZİLERİ KULLANARAK SES KAYNAĞI KONUMLANDIRMA

Yapay zeka teknolojilerinde son yıllardaki gerçekleşen gelişmeler ile birlikte ses kaydetme ve konuşma işleme uygulamalarında önemli bir artış gözleniyor. Ses kaydının kalitesini artırmak ve sesden ek konum ile ilgili bilgi edinmek için birden fazla mikrofonun geometrik olarak diziminden oluşan mikrofon dizileri son on yılda yaygın olarak kullanılmaktadır, ancak mikrofon dizileriyle ilgilenmek, dijitalden analoga dönüşüm yapmak için gelişmiş devre ve senkronizasyon ve yüksek kaliteyi sağlamak için eşleştirme devreleri gerektiriyordu. MEMS Dijital Mikrofonlarla ilgili son gelişmeler sayesinde, artık piyasada çok küçük, çok düşük-güç tüketimi olan ve yüksek kalitede ses alan mikrofon cihazları bulunuyor. Bu tezde, ses kaynağı lokalizasyonunda ve ses demeti oluşturma uygulamalarında kullanılan bir mikrofon dizisi oluşturmak için en güncel MEMS dijital mikrofon alternatifleri incelenmiştir. Farklı üreticiler tarafından kullanılan I2S ve TDM dahil alternatif dijital iletişim arayüzlerini test edildi ve sonrasında 4 mikrofonunu içeren kare şeklinde bir PCB üretilerek üzerindeki mikrofon dizisi konuşma tanıma ve telekonferans odası sistemleri de dahil olmak üzere gerçek dünya ortamlarında sağlamlığı ve hızlı performansı sebebiyle en yaygın kullanılan model olan ses gecikmeli tahmin modelini kullanarak değerlendirildi. Her bir kanalın gecikmesini doğru bir şekilde hesaplamak için, GCC-PHAT çapraz korelasyon yöntemi kullanıldı ve daha sonra dizi geometrisi girdileri kullanılarak ses varış açısı elde edildi. Son olarak, akustik port dış tasarımının olası olumsuz etkilerini ortadan kaldırmak için bir cihaz muhafaza kutusu tasarlandı ve üç boyutlu olarak üretildi.

TABLE OF CONTENTS

ACKNOWLEDGMENTS iii

ABSTRACT iv

ÖZET v

TABLE OF CONTENTS vi

LIST OF FIGURES viii

LIST OF TABLES xii

LIST OF SYMBOLS xiii

LIST OF ACRONYMS/ABBREVIATIONS xiv

1. INTRODUCTION 1

1.1. Capacitive Mems Microphones. 3

1.1.1. Digital and Analog Capacitive MEMS Microphones 5

1.2. MEMS Microphone Arrays 7

1.2.1. Broadside Microphone Arrays 8

1.2.1. Endfire Microphone Arrays 10

1.2.3. Custom Geometry Microphone Arrays 12

1.2.4. Commercial Microphone Arrays in Market. 14

1.3. Sound Source Localization Methods 15

1.3.2. Time Delay of Arrival using Cross-Correlation 15

1.3.1. Delay and Sum Beamforming 17

1.4. Thesis Outline 19

2. DIGITAL SOUND INTERFACES 20

2.1. Pulse Code Modulation (PCM) 20

2.2. Pulse Density Modulation (PDM) 21
2.3. Inter-IC-Sound (I2S) 23
2.4. Time Division Multiplexing (TDM) 25
2.5. Supported Interfaces by Platforms and Microcontrollers 28
2.6. A Workaround: Adding TDM Support to I2S Interface 30
3. DESIGN AND EXPERIMENTS 32
3.1. Evaluation of Microphone Alternatives 32
3.1. Tests with Evaluation Boards 34
3.2. PCB Design and Production 38
3.2.1. Schematics and Layout Design 39
3.2.1. PCB Production and Assembly 40
3.3. Sound Source Localization 42
3.3.1. Time Delay of Arrival using Cross-Correlation 44
3.3.2. Direction of Arrival Estimation 47
3.3.3. Beamforming Using Time Delay of Arrival Information 49
4. CONCLUSION 51
REFERENCES 53
APPENDIX A: DEVICE AND EXPERIMENT PICTURES 56
APPENDIX B: COPYRIGHT AND TRADEMARK INFORMATION 59

LIST OF FIGURES

Figure 1.1 Global Mems Microphone Market Size, 2010-2019 [2] 1

Figure 1.2 Horizontal cross-section view of a bottom port MEMS digital microphone [5]. 3

Figure 1.3 Horizontal cross-section view of two MEMS microphone designs: (left) a top port, chip and wire bonding design, and (right) bottom port flip-chip bonding design. 4

Figure 1.4 Images of two microphone diaphragms belong to an early MEMS microphone SP0103BE3 of Knowles, and improved design with larger area of Knowles KMM1 device [9]. 4

Figure 1.5 Blog Diagram of a typical analog capacitive MEMS Microphone. 5

Figure 1.6 Blog Diagram of a digital capacitive MEMS Microphone. 6

Figure 1.7 Signal-to-Noise improvement with sample signal waveforms. 8

Figure 1.8 Example geometry of a broadside microphone array (M1, M2, M3).. 9

Figure 1.9 Signal delay and beam forming illustration of a broadside array. 10

Figure 1.10 Example geometry of an endfire microphone array. 11

Figure 1.11 Signal delay and beam forming illustration of an endfire array.. 12

Figure 1.12 Example geometry of a custom microphone array (M1-M5) 3D view at left and top view at right. 13

Figure 1.13 Google Home (InvenSense INP621) and Amazon Echo (Knowles S1053-0090 V6) [16]. 14

Figure 1.14 Respeaker 4 Mic Array by Seeed Studio (SPU0414HR5H-SB) [17].. 15

Figure 1.15 Time Delay of Arrival Method Illustration with 2-Microphone Array.. . . . 16

Figure 1.16 Rotating Sound Source Illustration with 4-Microphone Array. 17

Figure 2.1 4-bit PCM quantization representation (blue dots) of a reference signal (red). 20

Figure 2.2 Internal block diagram of simple Delta-Sigma modulator [22].. 21

Figure 2.3 1-bit PDM modulation representation of a reference signal.. 22

Figure 2.4 Typical block diagram of I2S interface including one MEMS microphone. . 23

Figure 2.5 Typical block diagram of I2S interface including two channels [25]. 24

Figure 2.6 Internal block diagram of ICS-52000 TDM MEMS Microphone [27].. . . . 25

Figure 2.7 Block diagram of ICS-52000 TDM MEMS Microphone Array [27] 26

Figure 2.8 Timing diagram of ICS-52000 TDM MEMS Microphone Array [27].. . . . 27

Figure 2.9 Output data format of ICS-52000 TDM MEMS Microphone. 28

Figure 2.10 Timing diagram and parameters in I2S interface of BCM 2837 [28] 30

Figure 2.11 Block and Timing diagram of I2S to TDM conversion for 4-Mic Array.. . . 31

Figure 3.1 Block diagram of I2S interface between ICS43432 and Raspberry Pi Zero . 32

Figure 3.2 Sample stereo recording of 1000 Hz Sin Wave at 44.1Hz using ICS43432. . 33

Figure 3.3 Schematics and PCB of AUD-ICS52000 evaluation board [41]. 35

Figure 3.4 Figure 3.4. CLK (Yellow) and SD line (Green) on left and resulting noise for sinusoidal x1000 Hz tone recording using AUD-ICS52000 evaluation board. 36

Figure 3.5 Soldering of bare single ICS-52000 microphone (on the left) and successfully recorded human voice pulses using Teensy 3.2 (on the right). 37

Figure 3.6 New ordered ICS-52000 microphone (520QBA18Y) on the left and ICS-52000 on AUD-ICS52000 evaluation board (520PZ5355) on the right. 37

Figure 3.7 Prototype array build on top of epoxy board consisting new ICS-5200. microphones connected to Teensy 3.2 board.. 38

Figure 3.8 Spider Distribution Network [42] and implemented prototype. 38

Figure 3.9 3D Dimensions and PCB Land Pattern Layout of ICS-52000 [27].. 39

Figure 3.10 Schematics of Designed Microphone Array Device.. 40

Figure 3.11 2-Layer PCB Layout and designed PCB placement in 3D device box.. 41

Figure 3.12 Assembled PCB of Designed 4-Mic Array Device. 42

Figure 3.13 Teller Desk with Microphone Array Scenario Representation. 43

Figure 3.14 Voice spectra of males, females and children (1/3 octave) by effort [44]. . . 44

Figure 3.15 Direction of arrival experiment, top view. 44

Figure 3.16 Recorded data at 44.1 kHz for direction of arrival analysis.. 45

Figure 3.17 Voice activity detection. 45

Figure 3.18 Cross-correlation arrays for different 200-sample windows for same DoA. . 46

Figure 3.19 Time delay calculation between microphone pairs. 47

Figure 3.20 Calculated and actual direction of arrival. 48

Figure 3.21 Original and Corrected (delayed) waveforms. 49

Figure 3.22 Addition of original signals and Beamforming result. 50

Figure A.1 Experiment setup including ICS-52000 array with Raspberry Pi 3b+. 58

Figure A.2 Experiment setup including Teensy 3.2 and prototype device. 58

Figure A.3 Designed and produced PCB. 59

Figure A.4 Assembled 3D printed device box. 59

Figure A.5 Real-time recording of sinusoidal test waveform. 60

LIST OF TABLES

Table 2.1 PDM, I2S and TDM support of platforms and microcontrollers [28-34]. . . 29

Table 3.1 Comparison of reviewed and tested MEMS microphones [27][36][39][40].. 34

LIST OF SYMBOLS

C_d	Cross correlation of two signals collected at different microphone
cm	Centi-meters, metric unit
f_s	Sampling frequency
h	Height of sound source relative to microphone array
mm	Milli-meters, metric unit
r	Radius of the circle in which the sound source was rotated
$X_{34}[\theta, n]$	Resulting signal after beamforming of signal x_3 and x_4
Δ	Time delta / time delay
θ	Theta, direction of arrival angle
θ_{source}	Angle of the most possible sound source
τ	Resulting time delay
τ_{max}	Maximum time delay
∑	Signal summation
°	Degree, angle unit
Ω	Ohm, resistance unit

LIST OF ACRONYMS/ABBREVIATIONS

MEMS	Micro Electro-Mechanical Systems
3D	Three Dimensional
PCB	Printed Circuit Board
I2S	Inter-IC Sound
TDM	Time-Division Multiplexing
PDM	Pulse-Density Modulation
TDM	Time-Division Multiplexing
PCM	Pulse-Coded Modulation
GCC-PHAT	Generalized Cross Correlation with Phase Transform
ECM	Electret Condenser Microphones
DOA	Direction of Arrival
TDE	Time Delay Estimation
SNR	Signal to Noise Ratio
MIC	Microphone
ALSA	Advanced Linux Sound Architecture
DOA	Direction of Arrival
ASIC	Application Specific Integrated Circuit

I/O	Input/output
AI	Artificial Intelligence
ADC	Analog to Digital Convertor
DAC	Digital to Analog Convertor

1. INTRODUCTION

Today almost every mobile device including smartphones, tablets, laptops and digital cameras utilizes MEMS microphones instead of old-fashioned Electret Condenser Microphones (ECM) developed in early 1960s at Bell Labs [1]. Very similar to operating principle of ECM microphones, most of MEMS microphones use capacitance change between parallel diaphragms thanks to vibration of incoming sound wave. There are also some other MEMS microphone technologies such as piezoelectric microphones which are transferring audio mechanical energy to electric energy directly. During this study, we mostly focused and used capacitive MEMS microphones. Thanks to higher SNR ratios, small sizes and low-power consumption, MEMS microphones are widely used as sensors for both recording sound and detecting various environment changes.

According to IHS Markit, London-based global information provider, number of MEMS microphones in smartphones are increasing up to four or five MEMS microphones per single device as a result of need for better sound quality in hands-free calling, clearer voice commands for AI Assistants, noise cancellation in video recording, and enhanced call and sound recording performance [2]. Figure 1.1 below shows MEMS microphone market growth since 2010.

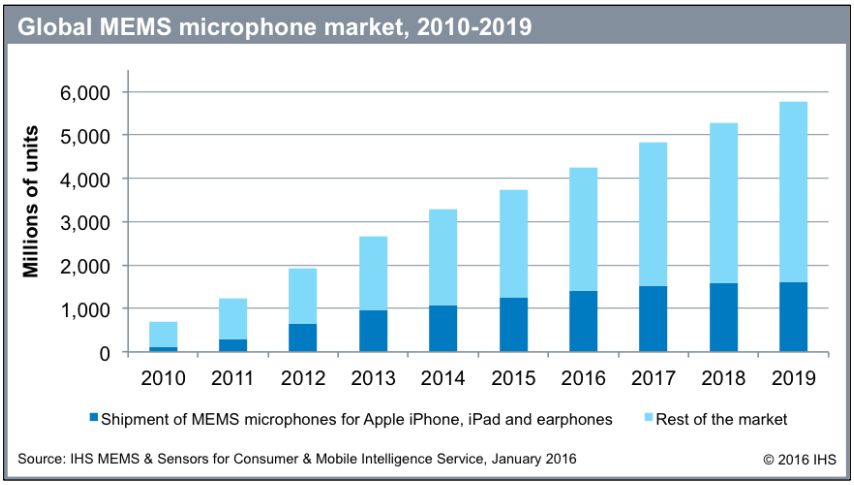

Figure 1.1. Global Mems Microphone Market Size, 2010-2019 [2].

Another technological area which needs better sound quality and additional information about acoustic sources is modern meeting rooms which are allowing local and remote participants to meet

and collaborate with both sound and display sharing experience. In these systems, participants need to hear each other and see the person who is speaking in the display of camera, using location and angle information obtained using acoustic sound sources. Very similar localization needs also exist in radar applications, search and rescue, medical treatment and other areas. In order to estimate the direction of arrival (DOA) of an acoustic sound source, time delay estimation (TDE) algorithms with a microphone array is used widely and data coming from different microphone channels were compared with each other by using signal comparing functions to find peak sound signal appeared on correct angle of direction.

Using theoretical knowledge of waves and known speed of sound wave in the air, we know that sound signals arrive at the microphones at different times when a microphone array is formed to place microphones in geometrically different places. It is aimed to determine the location of the sound source by examining the sound signals taken from different microphones. At this point, omni-directional and identical pickup response of the microphones in array is very critical to eliminate possible estimation mistakes. Since MEMS microphones have an omnidirectional pickup response by their mechanical design, which will be mentioned in following section, they respond equally to sounds coming from any direction and it is easier to form an array of multiple MEMS microphones to obtain a directional response or a beam pattern. To increase sensitivity to sound waves coming from one direction/angle compared to other directions, preferably to catch signals only from sound source which is aimed to be recorded, beamforming microphone arrays can be designed and used.

Microphone array beamforming is an inter-disciplinary topic including both mechanical and electrical design and digital signal processing parts. Before going through delay calculation between each microphone in array geometry, ensuring all of channels are containing accurate sound data with adequate level of SNR and all channels are synchronized using proper micro-processor and communication protocols are crucial. To obtain best performance to get sound data from multiple microphone, MEMS microphone alternatives from different manufacturers and different communication methods will be studied and explained in following sections.

While the MEMS microphones usually show very flat response in their operating range, external device design and location of the microphones inside a box of device case affects this response. In order to minimize these side effects on microphone performance, MEMS microphone manufacturers have mounting recommendations for their products including acoustic port height, diameter and cavity.

1.1. Capacitive MEMS Microphones

The utilization of technological developments in MEMS (micro electro-mechanical systems) area for microphones has accelerated the development of smaller microphones with very high performance. In the last decade, electret condenser microphones replaced with MEMS options thanks to small size, lower power consumption, resistance against soldering heat and a very high SNR and sound quality [3, 4]. They are also offered in very small packages that are completely compatible with surface mount assembly processes and resistant to high-temperature soldering thanks to excellent temperature characteristics. Because of their digital interface, they are not as much of vulnerable to picking up RF interference with the interconnect wires on the printed circuit board which could cause noise or distortion of the sound signal [4]. So biggest smartphone manufacturers now have more options to place them in their products even very close to the cellular, Bluetooth and Wi-Fi antennas.

Condenser silicon MEMS microphones are basically acoustic sensors which are produced using automated semiconductor fabrication methods on silicon wafers. Thin layers of various materials are deposited on top of silicon wafer and unwanted parts are then removed using etching techniques. Figure 1.2 shows cross-section view of a example bottom port MEMS microphone [5]. They mostly include two chips between metal cap and ceramic substrate. Firstly, an acoustic sensor is responsible for conversion of acoustic wave coming from sound hole into an electrical signal. This part is fabricated by creating a diaphragm as shown as black line in Figure 1.2.

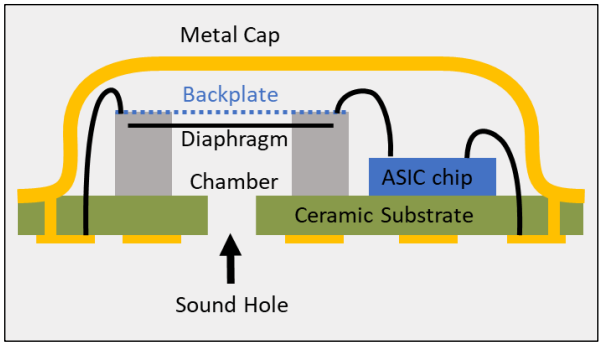

Figure 1.2. Horizontal cross-section view of a bottom port MEMS digital microphone [5].

There are some other packaging designs including top port and bottom port with flip-chip bonding. G. Feiertag [6], designed and described a microphone using flip chip technology in MEMS

packaging, instead of chip and wire bonding, the transducer and an ASIC are bonded on a ceramic substrate on top of microphone area in metal cap as shown in Figure 1.3. There are some other reference studies showed that sensor chip with two backplates leads to a higher sensitivity and a higher SNR [7][8].

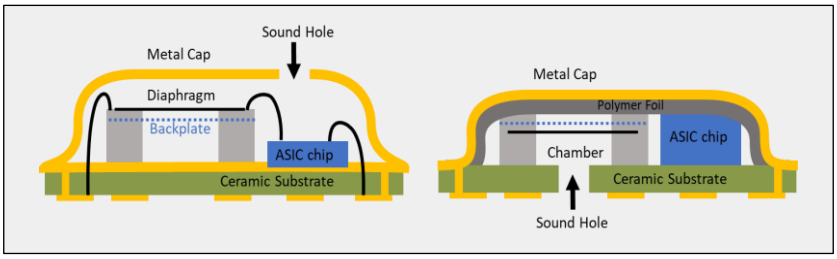

Figure 1.3. Horizontal cross-section view of two MEMS microphone designs: (left) a top port, chip and wire bonding design, and (right) bottom port flip-chip bonding design.

For all designs mentioned, fluctuations in air pressure shaped by acoustic waves, bends thin diaphragm whereas the back-plate preserves its fixed position while the acoustic waves passes through its holes. This movement creates a capacitance change between the backplate and the diaphragm, and this capacitance is transferred to ASIC (application-specific-integrated-circuit) using thin wires. Figure 1.3 shows images of an example microphone diagrams which is belong to an early two MEMS microphone by Knowles [9]. In the left image the microphone diaphragm occupied only 13% of the die area whereas improved device on the right, Knowles KMM1, has two active diaphragms in nearly 35% of the die area, while total area of device is reducing. Today we have devices including four and more diaphragms with larger area ratios.

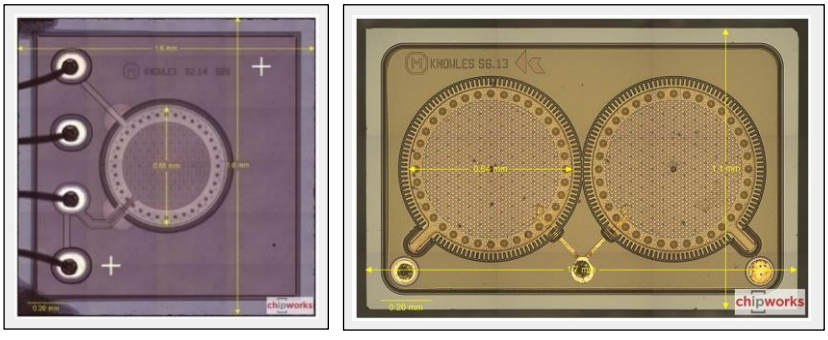

Figure 1.4. Images of two microphone diaphragms belong to an early MEMS microphone, SP0103BE3 of Knowles, and improved design with larger area of Knowles KMM1 device [9].

Capacitance difference created by diaphragm and backplate is converted into an electrical signal via the ASIC (application specific integrated circuit). A charge pump in ASIC places a fixed charge on the microphone diaphragm and then measures the voltage differences by capacitance with the solid backplate [10]. Then it generates an output voltage that is proportional to the instantaneous air pressure level. ASIC is also responsible for signal amplification, transmission to the external systems. Thanks to developments in MEMS technology in last decade, ASIC chips inside MEMS Microphones may include more advanced circuitries to fulfill functions of external devices including analog to digital conversion, frequency band filtering and audio codec operations.

1.1.1. Digital and Analog Capacitive MEMS Microphones

There are two groups of MEMS microphones based on the type of electrical output signal and transmission of information from chip to external circuitry. When ASIC chip shown in Figure 1.2 includes only amplifier and transmits sound signals as analog magnitude data to the outer system, this type of microphones is called analog MEMS microphones. Block diagram of a typical analog MEMS microphone is given in Figure 1.4. Analog microphones typically two input pins including the VDD and Ground and 1 output data pin. The interface required for analog microphones has relatively simple compounds but handling of the analog signal needs careful design of the PCB and wires to evade catching noise during the transmission of signal between the analog output and input of the integrated circuit getting analog signal. Usually, an audio ADC with low noise is also required to translate the analog output of microphones to binary signals for digital processing [10].

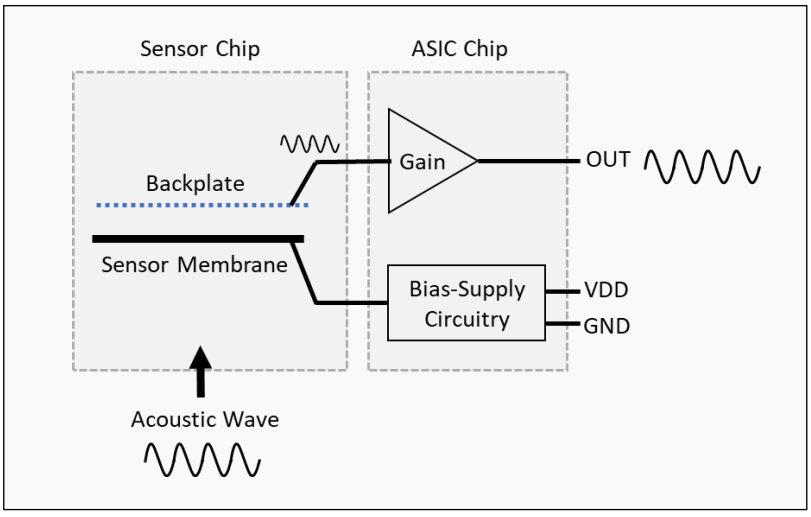

Figure 1.5. Blog Diagram of a typical analog capacitive MEMS Microphone.

When ASIC chip includes additional sub circuitries for the analog to digital conversion, filtering, and other codec operations, this type of microphones is called as digital MEMS microphones and they have digital outputs changing between logic zero and logic one. Figure 1.5 below shows a typical digital MEMS microphone block diagram.

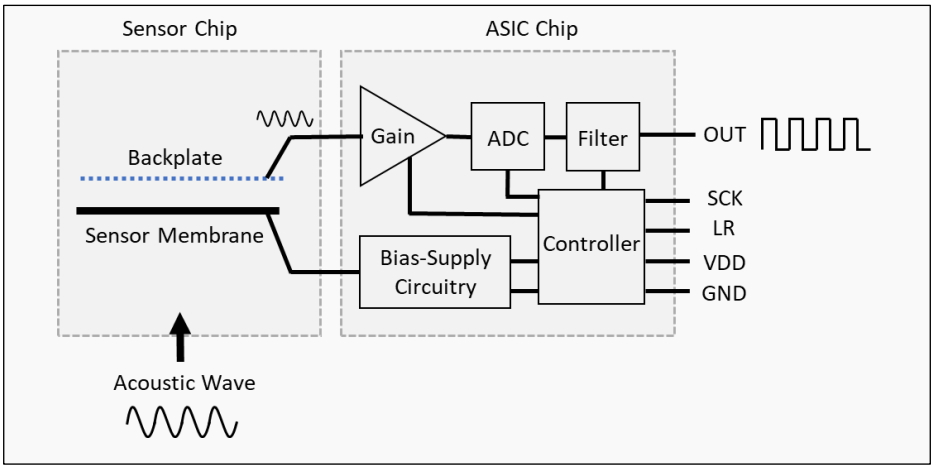

Figure 1.6. Blog Diagram of a digital capacitive MEMS Microphone.

Most of digital MEMS microphones have some additional input pins compared to analog microphones including a SCK (clock) and a L/R (Left/Right) control pins. Since at least two additional pads are essential for digital microphones, analog microphones have a smaller footprint and functions combined in the ASIC chip increases the need for a larger ASIC.

SCK (Clock) input is used for different purposes according to output data format. If it is a microphone with digital pulse density modulation (PDM) output, clock signal drives delta-sigma modulator which is translating the analog sound signal to a binary pulse density PDM signal. If microphone has Inter-IC Sound (I2S) output data format, clock signal is used to drive analog to digital convertor which creates I2S data format output synchronized with input clock signal. Most of the early digital microphones were using PDM format that is including an oversampled single-bit data stream and number of pulses in unit duration is proportional to the instantaneous air pressure level but today, demand for I2S and TDM audio interfaces is also growing because of their some advantages for applications that are mostly digital [11].

Although codecs or other post processing steps seem not essential for digital MEMS microphones, usually the PDM output must be converted to multi-bit pulse code modulation (PCM)

format from single bit PDM. This step is generally done by codecs and SoCs that have filters converting PDM format into PCM format or software filters implemented in Microcontroller firmware.

Another important concern related to analog MEMS microphones especially in multi-microphone systems or arrays is ensuring the matching between microphones for both time delay and signal distortion. Since all signals from different microphones follow different paths before reaching an analog to digital convertor, noise and time delay of asynchronous signals are hard to distinguish.

The digital sound signals generated by digital MEMS microphones are more resistant to noise compared to analog signals as a result of the binary transmission of voltage, but there is still important concerns to consider like parasitic capacitance distortion, inductance and impedance mismatches between the output of microphone and external micro controller especially when application design is requiring long travel distances of signals between microphone and controller units. In order to minimize possible distortion while using digital MEMS microphones, various digital output data format alternatives and tradeoffs will be described in Digital Sound Interfaces section in detail.

1.2. MEMS Microphone Arrays

Most of MEMS microphones are omnidirectional which means that they capture sound signals from any direction at equal sensitivity although there are some directional MEMS microphone exceptions [12][13]. As a natural result of this omnidirectional characteristic, captured acoustic signals include unwanted noise besides desired sound signals. To diminish unwanted noise and increase quality of sound being collected with omnidirectional microphones, MEMS Microphone arrays are utilized widely. These arrays contain multiple microphones in a 2D or 3D geometry and offer better sensitivity and higher SNR (signal-to-noise ratios) when two or more sound signals are used to analyze direction of arrival and calculate directional response using beamforming techniques.

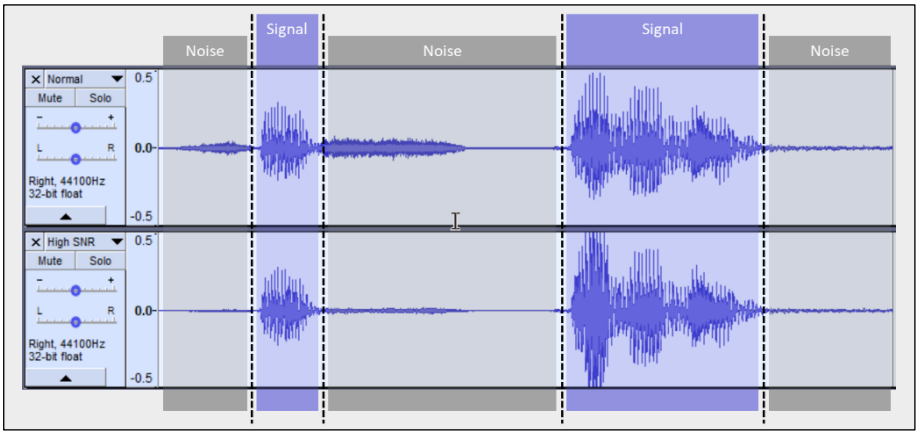

Figure 1.7. Signal-to-Noise improvement with sample signal waveforms.
(Normal signal at top and High SNR Signal improved using beamforming at bottom)

Before combining sound signals from each element of an array, mechanical to electrical conversion and following processes like amplification and filtering are done by electronic circuitry. For MEMS digital microphone case, these operations are done by ASIC chip inside same MEMS device and output of each microphone includes an improved signal. When it comes to combining signals from each microphone, to ensure that improvements are preserved and increased, it is necessary to match performance and characteristics of each microphones. Most important characteristic to be matched is the sensitivity of the microphones. Thanks to infinitesimal variation in batch processing used in semiconductor manufacturing processes, MEMS digital microphones are highly matched and most suitable option for microphone arrays.

There are different categories of MEMS microphone arrays, based on application types, their array geometry and their working principles.

1.2.1. Broadside Microphone Arrays

In the most of sound capturing applications, it is wanted to ignore undesired sounds coming from minor sound sources surrounding the actual sound source which is desired to be captured. Undesired sound sources are referred as noise and signal-to-noise ratio of sound signals can be improved by detecting and removing signals coming from unfocused directions. Broadside microphone arrays are one of the common types of arrays and it is formed by 1D or 2D placement

of microphones as an array where line or plane consisting array of microphones is perpendicular to the sound source desired to captured [14].

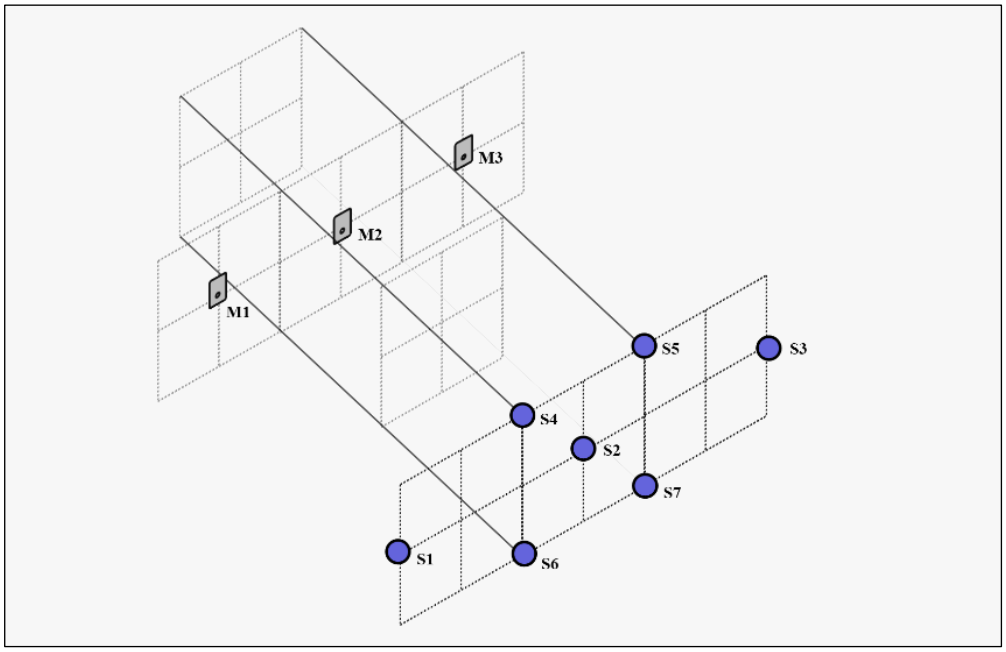

Figure 1.8. Example geometry of a broadside microphone array (M1, M2, M3).

With broadside array geometry as shown in Figure 1.7., sound wave from perpendicular source (Source 2) reaches microphone at the center (M2) first, and far microphones (M1 and M3) later based on displacement between microphones. When signals captured synchronically form each microphone are summed, signal coming from Source 2 strengthen constructively whereas signals coming from other directions are weaken due to the fact that each microphone receives sound waves with different delays according to speed of sound and distance.

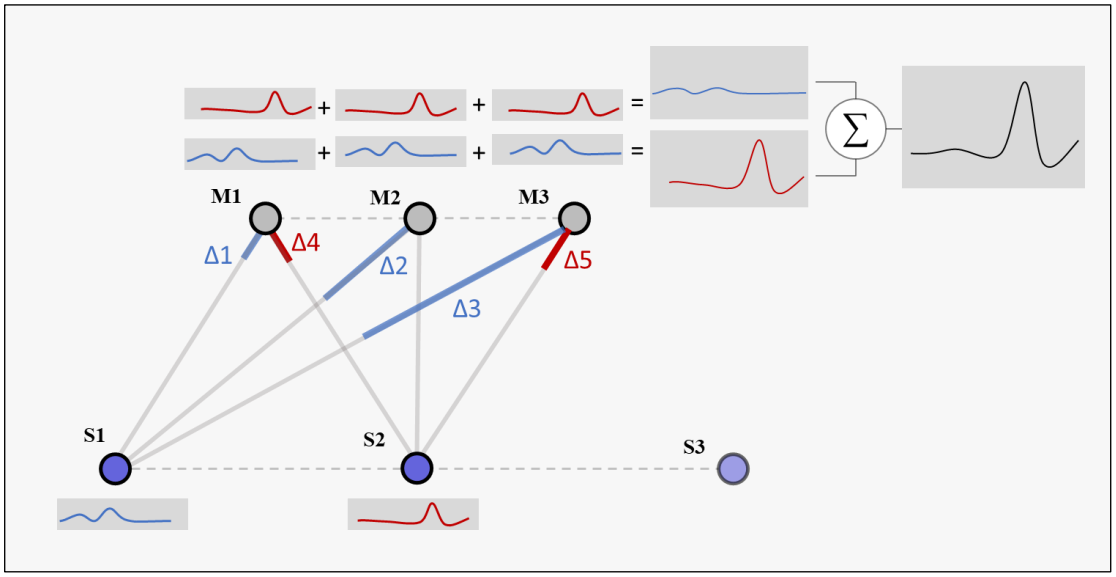

Figure 1.9. Signal delay and beam forming illustration of a broadside array.

Delays due to the different distances (Δ1,2,3,4,5) and cancelation of signals from side sound sources (Source 4 in figure) are represented in Figure 1.8. Red pulse coming from Source 2 arrives microphones with smaller delays (Δ4 and Δ5) whereas blue pulse coming from Source 1 arrives microphones with larger delays (Δ1, Δ2 and Δ3). As a result of this large phase difference, red pulse is amplified while blue pulse is attenuated during summation operation.

Laptop microphones optimized for video calls located on top corners of screen are examples of application area for broadside arrays which allow amplification of sound only from person in front of computer.

1.2.2. Endfire Microphone Arrays

Another common type of microphone array based on geometry is endfire microphone arrays which is formed by placement of microphones as an array where line or plane consisting array of microphones is passing through to the sound source desired to captured.

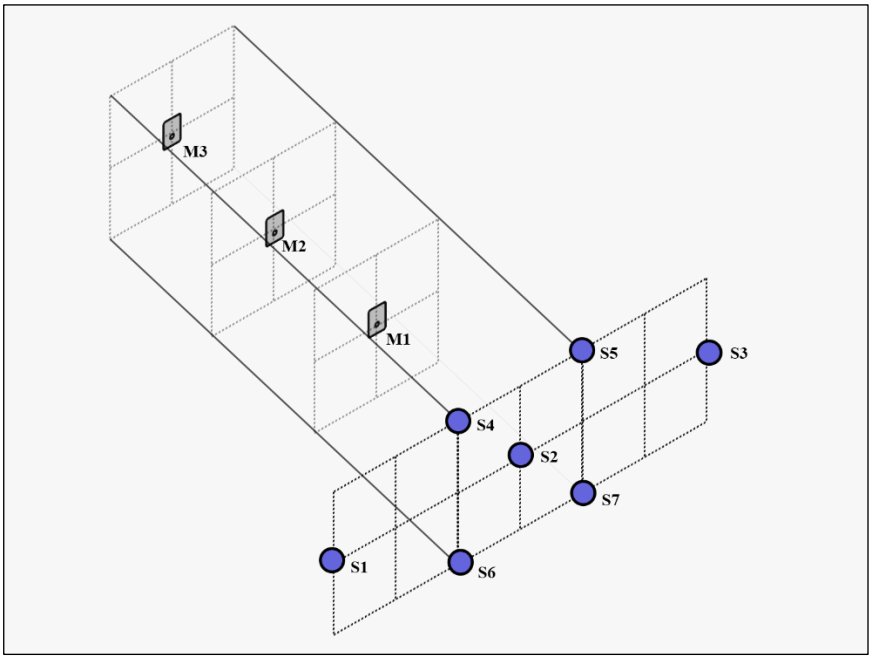

Figure 1.10. Example geometry of an endfire microphone array.

With endfire array geometry as shown in Figure 1.9., sound wave from perpendicular source (S2) reaches microphone at front (M1) first and far microphones (M2, M3) later based on displacement between microphones. Sound signals are captured synchronically form each microphone and delayed according to time delay calculated by [known distance between microphones] / [speed of sound] before they were summed. When delayed signals are summed, source 2 is strengthen constructively whereas signals coming from other directions are weakening due to the fact that each microphone receives sound waves with different delays according to speed of sound and distance. Very similar to the broadside arrays, endfire microphone arrays are adding sound signals from focused direction in a constructive way whereas signals coming from other sources (S1, S3, S4, S5, S6 and S7) weaken due to fixed delay applied which is calculated as maximum delay among microphones.

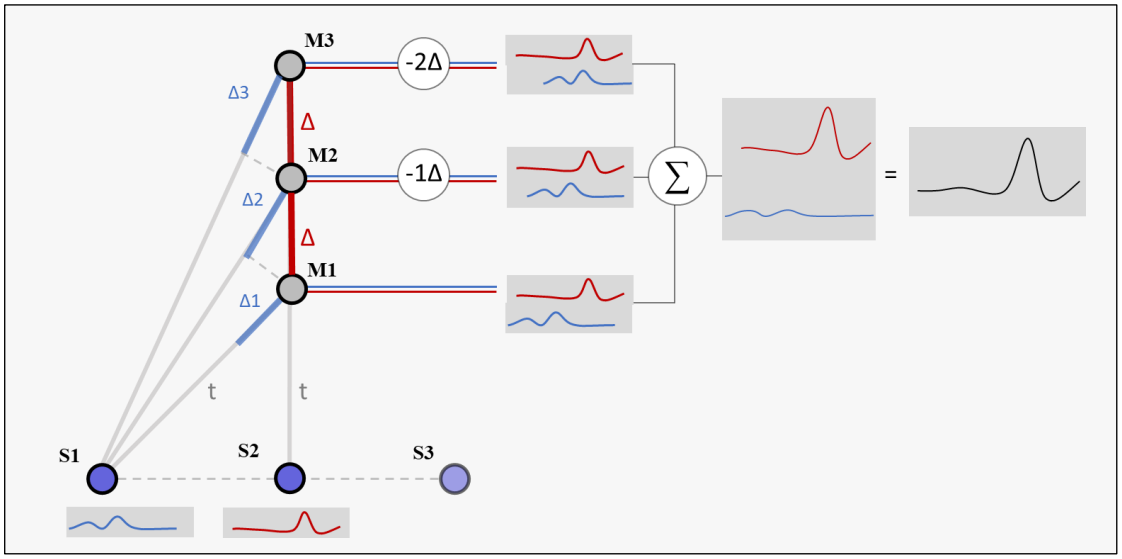

Figure 1.11. Signal delay and beam forming illustration of an endfire array.

Simple working principle of an endfire microphone array is represented in Figure 1.10. Signal from S1 source reaches M1 with Δ1 delay comparing with S2 sound source. From Microphone M2 point of view, sound S1 is delayed by Δ1+Δ2 whereas sound S2 is delayed by Δ. Since signal exiting M2 microphone is forwarded by -1Δ with electronic circuitry, red signal which is desired to be recorded is corrected by +Δ - Δ = 0 operation whereas blue signal from S1 sound source is shifted by Δ1 + Δ2 - Δ which is not equal to zero. Due to this distortion among microphones blue signal is weakened after summation operation while signal from S2 source is amplified without disturbing by summation of same phase signals from each microphone.

1.2.3. Custom Geometry Microphone Arrays

As it is explained above for both types of geometries, array implementation enhances sound quality of desired sound source whereas attenuating other noise sources. Most important difference between them is directional sensitivity of endfire arrays. Broadside arrays are not capable of capturing signals from only one direction because of both its geometry and omnidirectional MEMS microphones that make it impossible to distinguish direction. On the other hand, one possible disadvantage of endfire array is attenuation of side signals and it is not at the order of broadside

arrays with similar size. In order to eliminate individual disadvantages of both geometry, complex array geometries including both geometries in single array can be used.

An example design of custom geometry microphone array is Figure 1.11 given below shows an example design of custom geometry microphone array. It includes four microphones horizontally located on corners of a square and one located at center of square where plane consisting microphones is perpendicular to possible locations of sound sources.

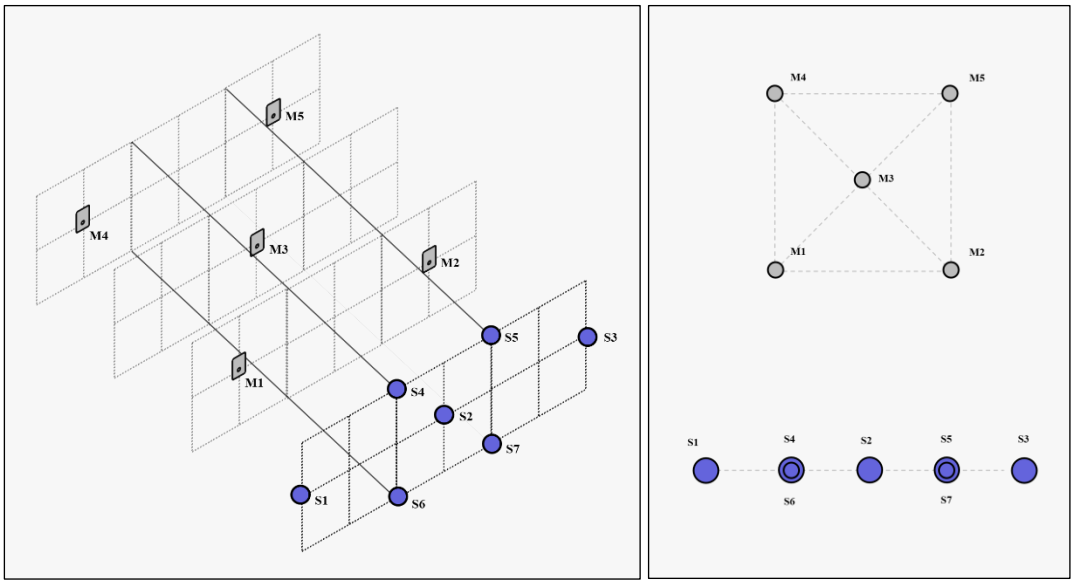

Figure 1.12. Example geometry of a custom microphone array (M1-M5)
3D view at left and top view at right.

Direction of arrival estimation in this type of custom geometries includes more calculation methods since some pairs act as broadside whereas some microphone pairs are act as end-fire array for specific sound source locations. Instead of making calculation for each pair of microphones, two far-most microphone pairs are chosen first to understand region of sound source roughly (M1-M5 and M2-M4 for Figure 1.11) and then most broadside pair is used to calculate accurate position of dominant sound source. Complete calculation steps for similar array geometry will be discussed in experiment section.

1.2.4. Commercial Microphone Arrays in Market

There are a few commercial products on the market using MEMS mic arrays for their voice assistant devices. The first one, Amazon Echo has 7 MEMS analog microphones (SPA1687LR5H-1 from Knowles). Contrary to it, Google Home contains only two MEMS microphones (InvenSense INMP621). These microphones have omnidirectional pickup response from its bottom port with low power consumption and very low noise, which enable capture of sound in loud environments according to InvenSense Inc [15]. The MEMS microphones also have extended wideband frequency response which helps in producing natural sound with the ability to hear commands clearly. There is no quantitative comparison of sound quality they capture but Google states that its 2 mics are capable of low noise and high-quality recordings better than Amazon's 7 mics array.

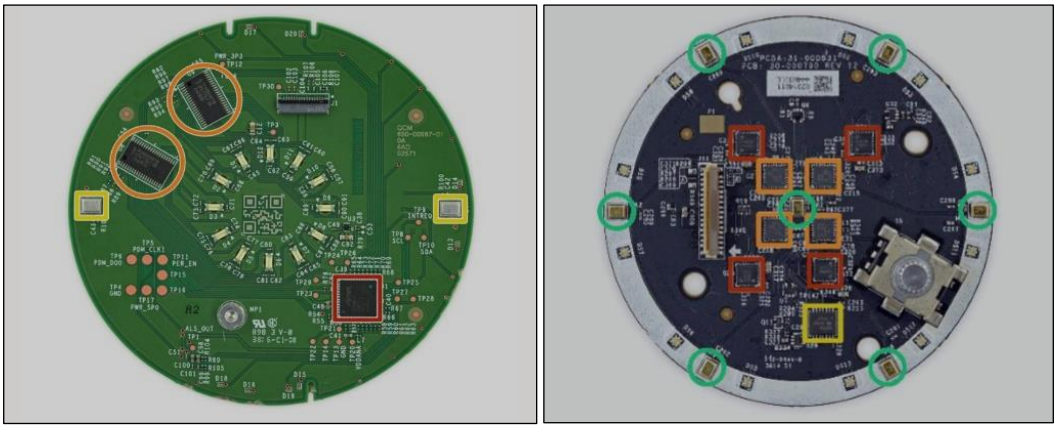

Figure 1.13. Google Home (InvenSense INP621) and Amazon Echo (Knowles S1053 0090 V6) [16].

There are some other array alternatives from different producers which are widely used for academic or hobby purpose experiments. ReSpeaker 4-Mic Array for Raspberry Pi is an easy to use microphone array designed for AI and voice applications and consisting four microphones at the corner of a square shape expansion board compatible with Raspberry Pi. As shown in Figure 1.13, Respeaker consist of Analog MEMS microphones (SPU0414HR5H-SB) and analog to digital convertor (AC108) to process analog data coming from 4 microphones.

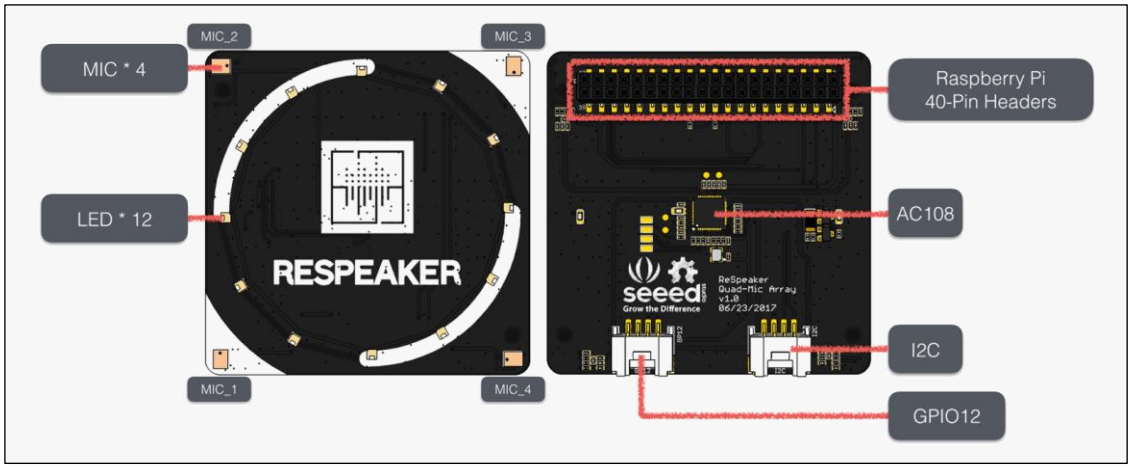

Figure 1.14. Respeaker 4 Mic Array by Seeed Studio (SPU0414HR5H-SB) [17].

1.3. Sound Source Localization Methods

For each type of array geometries mentioned in previous section, it is easy to eliminate sound signals from undesired directions when direction or location of desired sound source is known relative to microphone array location and position. In most real-life environments and sound sources are not fixed in 3D space and there are many different sources which are simultaneously generating sound signals with different characteristics. In these complex environments, first step before filtering out undesired signals is finding the position of desired sound signal. There are two most common methods in order to find desired sound source location or direction. The first is Time Delay of Arrival method and the second is Delay and Sum Method [18]. Working principles and drawbacks of each methods are discussed below.

1.3.1. Time Delay of Arrival using Cross-Correlation

For the simple array geometry shown in Figure 1.13, signal from S1 sound source arrives M2 microphone with time delay τ comparing with M1 microphone and angle θ can be calculated as $\theta = \arcsin(\tau \cdot c / d)$ where c is the speed of sound equal to 343 m/sec. Time delay calculation of signal pairs which are obtained from microphones M1 and M2 is a signal processing operation. To calculate

delay or shift between signals, cross-correlation calculation in time domain or frequency domain is utilized in time delay of arrival method.

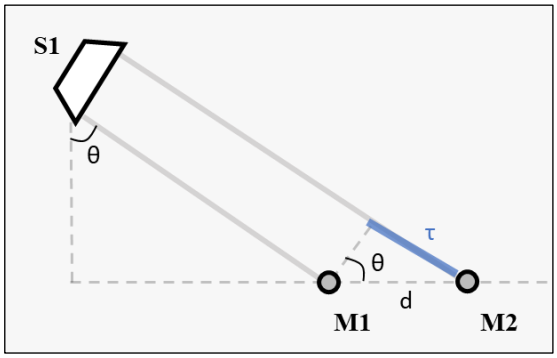

Figure 1.15. Time Delay of Arrival Method Illustration with 2-Microphone Array.

Cross-Correlation C_d of two signals x_1 and x_2 generated from same sound source S1 but collected at M1 and M2 microphones respectively can be calculated as following equation [19] where k is equal to possible values of time delay τ between microphones.

$$C_d[k] = \sum_{n=0}^{N} x_1[n].x_2[n+k] \tag{1.1}$$

For Figure 1.13, maximum time delay τ is equal to known distance d between microphones divided by 343 m/sec speed of sound. Therefore, C_d equation above is evaluated with k values spanning from zero (sound source in midline of microphones) to $\pm\tau_{max}$ = d/343 (sound source in line containing microphones) to find point with maximum correlation. Resulting time delay τ for desired signal between two signals is then calculated as following equation where f_s is sampling frequency of signals.

$$\tau = argmax(C_d[k]) \,/\, f_s \tag{1.2}$$

The first step of calculating $C_d[k]$ for different k values requires large computing power especially signal count and sample size is getting larger. In order to accelerate this analysis, time domain signals can be converted to frequency domain and calculations can be simplified on two steps: Spectral Power Density of signals are obtained by multiplication of Fourier transform of signals and then inverse Fourier Transform of result to obtain cross-correlation.

Most important disadvantage of Time Delay of Arrival method is that it can only detects one dominant sound source when there are multiple sound sources need to be detected and separated. Another important consideration with this method is that possible adjacent noises on sample windows can cause miscalculated cross-correlations especially when microphone distance is very small comparing with sound wavelength.

1.3.2. Delay and Sum Beamforming

Delay and sum beamforming is very simple method, but it includes too much repetitive calculations to accurately guess actual direction of one or more sound sources. After scanning all possible locations of sound source by calculating resulting signal magnitudes for each delay and sum beamforming step, the angle gives maximum level of signal power is selected as the best guess of sound source location.

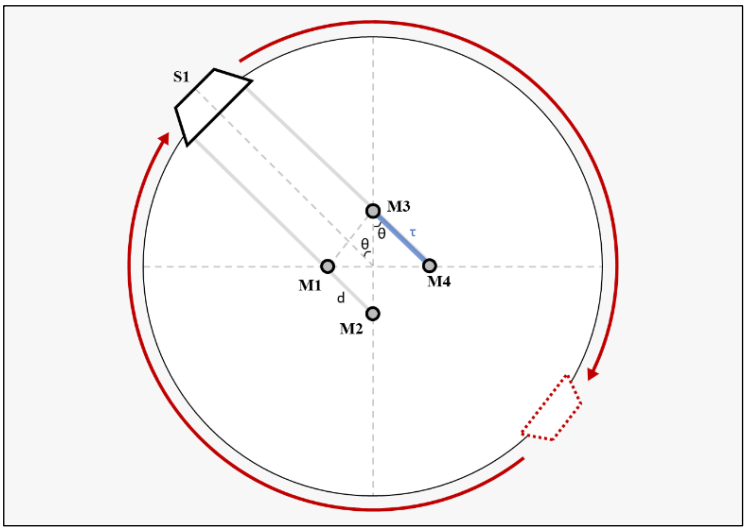

Figure 1.16. Rotating Sound Source Illustration with 4-Microphone Array.

For the simple array geometry given in Figure 1.14, when location of source S1 is unknown, time delay τ between for any angle on circle for each microphone pair is calculated and signals are summed after calculated delays are added to synchronize signals as equation below.

$$\tau_{34}[\theta] = \frac{\mathrm{d}\,.\cot(\theta)}{Speed\ of\ Sound} \tag{1.3}$$

$$X_{34}[\theta, n] = \mathrm{x}_3[n] + \mathrm{x}_4\left[n + \tau_{34}[\theta]\right] \tag{1.4}$$

$X_{34}[\theta, n]$ represents resulting signal after beamforming of signal x_3 and x_4 where n is the index of current sample. This calculation is done for each possible sound source angle and then, resulting beamformed signals are compared according to their magnitude and the largest one is estimated as possible sound source. Angle of most possible sound source θ_{source} is given as below equation.

$$\theta_{source} = argmax(X_{34}[\theta]) \tag{1.5}$$

Contrary to Time Delay calculation method, it is possible to detect more than one sound sources using delay and sum method. Angles with signal power larger than threshold power are recognized as different sound sources. Each sound source can be separated and beamformed using delay parameters obtained circular scanning of possible angles. Possible disadvantage of this method is high computing power needed for scanning each possible position.

For all of geometries and localization methods, matching of microphone hardware properties such as frequency response and sensitivity is crucial for accurate calculations. Any outer acoustic case or cavity adding extra path for specific sound signals can cause large mistakes in time delay and beamforming calculations.

1.4. Thesis Outline

Until this point in first chapter, analog and digital capacitive MEMS microphones and their working principles were discussed. Usage areas of microphone arrays were introduced, and common microphone array geometries were compared detailly. Advantages and drawbacks of different types of microphone arrays were discussed from sound source localization perspective. In the last part of this chapter, different sound source localization methods were discussed and evaluated including beamforming methods to amplify desired signals.

In Chapter 2, digital sound interfaces used by MEMS digital microphones including PDM (Pulse Density Modulation), I2S (Inter-IC Sound) and TDM (Time-Division Multiplexing) are evaluated and discussed. Improvements, new needs and limitations in microphone communication interfaces are discussed and compared along with support for different microcontrollers briefly.

In next Design and Experiments chapter, result of experiments using each Digital Sound Interface with products of different MEMS Microphone manufacturers are to be discussed firstly. Then microphone alternatives are to be evaluated at hardware level, and PCB design and hardware production steps are mentioned. At the last part of this chapter, sound source localization experiment and its outcomes are discussed detailly.

In Last Chapter, the document is concluded with a brief summary of the achieved. The results of the designed and produced microphone array and learnings during thesis work are explained. The future works of the project which improve microphone array device are also shared including effect of microphone positions, acoustic port and device case design.

2. DIGITAL SOUND INTERFACES

Aligned with technological developments around portable electronic devices and need for better sound quality for AI assistant space, digital audio interface between microphone and IC is important topic effecting circuit size, maximum sampling frequency, robustness and quality of recorded sound signals. As digital MEMS microphones are becoming more popular, new digital video interfaces which are using digital outputs and ensuring high speed of communication with minimum circuit resource are introduced in last decade. These new interfaces are mostly more cost-effective than analog communication since analog portions of a sound signal chain is remained only between input and output of sensors and everything outside is transmitted digitally [11]. In this chapter, most popular digital sound interfaces PCM, PDM, I2S, and TDM will be discussed and compared.

2.1. Pulse Code Modulation (PCM)

Pulse-code modulation (PCM) is a technique used to convert any analog signals to the digital bit arrays. It is the typical form of digital audio signals stored in a computer memory or transmitted to a microcontroller and preferred widely by digital audio systems. Amplitude of analog audio signal is sampled at a defined frequency and value of sample is quantized to nearest step value available in discreate output range. There are two important parameters effecting output's consistency to the original analog signal. The first one, the sampling rate is the quantity of samples taken in a second interval and the second one is bit-depth, which measures the quantity of discreate levels in output range. Figure 2.1 shows a typical PCM operation on a reference sin wave (red curve) where bit-depth is 4-bit which means 16 possible values in amplitude axis.

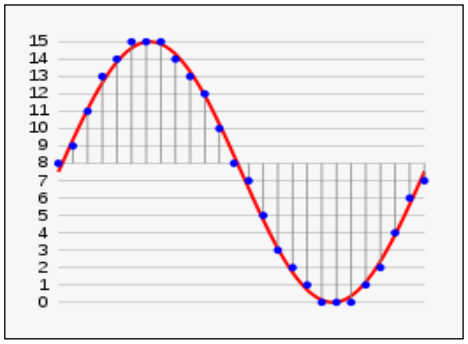

Figure 2.1. 4-bit PCM quantization representation (blue dots) of a reference signal (red).

Regarding possible information loss because of limited sampling of analog (continuous signal) to a discreate output during PCM operation, the Nyquist–Shannon sampling theorem demonstrates that it is possible to operate without distortions in intended frequency bands if sampling frequency is chosen at least twice that of the maximum frequency exist in the input audio signal [20]. Human voice frequency band starts from approximately 120 Hz and reaches to about 3800 Hz. In order to record and transmit voice signal without distortion, minimum sampling frequency needed is bigger than twice of 3800 Hz according to the Nyquist–Shannon sampling theorem, and 8k Hz is minimum industry standard for digital audio applications. Although sampling theorem ensures no loss or distortion on desired frequency band, if signal has inputs at frequency of Nyquist frequency (Sampling Frequency/2) or higher, it will add aliasing distortion to other signal parts below the Nyquist frequency and therefore resulting signal will not be correct representation of reference input signal.

Today, most common sampling frequencies are 48 kHz (DVD videos) and 44.1 kHz (Music Files) whereas there are some special sound devices offering higher frequencies on the order of 96 kHz and 192 kHz, but the further benefits as a result of higher sampling frequency is debated [21].

2.2. Pulse Density Modulation (PDM)

Pulse Density Modulation (PDM) is another type of modulation used to convert a continuous analog signal to a binary signal. Contrary to PCM, the relative density of the pulses corresponds to amplitude of reference signal, not the amplitude of pulses in signal is obtained and transmitted. The output of a PDM microphone is a single bit data stream working on high sample rate typically from 1 MHz to 3.25 MHz, and this single bit stream is generated in Delta-Sigma modulator working in the mic. The first step in Delta-Sigma (ΔΣ) modulation is delta modulation which the change in the signal (delta) is encoded. Figure 2.2 shows block diagram of simple Delta-Sigma modulator.

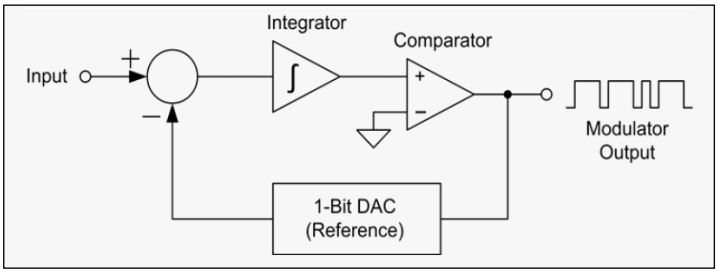

Figure 2.2. Internal block diagram of simple Delta-Sigma modulator [22].

Negative feedback loop consisting a 1-Bit DAC is adding sigma component to input signal before passing through integrator component. Resulting signal obtained from sinusoidal input signal is shown in Figure 2.3 below. Input signal (red) is converted to PDM output signal (blue discreate). High frequency pulses in output represents zero-crossings and flat regions represents peaks at sinusoidal input signal.

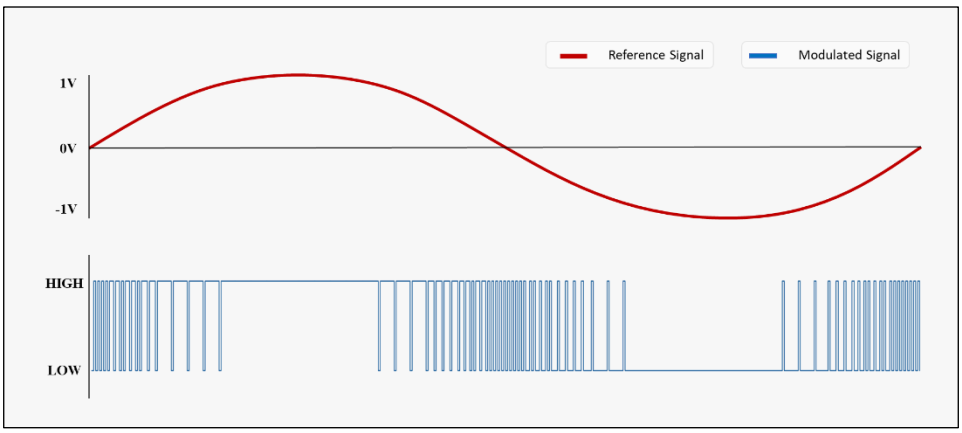

Figure 2.3. 1-bit PDM modulation representation of a reference signal.

Since PDM format includes only one-bit output, its transmission and processing are easier than PCM format and it brings the benefits of digital, including low noise and resistance to interfering signals, at low cost [23]. PDM microphones has become popular firstly in size-constrained mobile devices such as cellphones and small computers as a way to deliver audio from microphones to the signal processor efficiently. Clock and Data Input lines are two required lines to transmit one or two audio channels. Each slave device uses alternate edges of the clock, which is generated by master, to output their signals via common data line. The bandwidth of the sound signal increases as the frequency of clock pulses rises, therefore clocks with lower clock rates are required for systems which reduce of bandwidth valuable than decrease in the power consumption.

PDM's high sample rate is hard to be processed by standard microcontroller communication interfaces and decimation to a lower sample rates is needed in most practical applications. This operation is usually done by decimation filter placed in an audio codec or a digital signal processor placed between microphone and microcontroller. Output of decimation filter is at typically between 16 and 48 kHz.

2.3. Inter-IC Sound (I2S)

Inter-IC Sound (I2S or I^2S) is the most common digital audio interface intended for digital sound data transmission between integrated circuits and other audio processing components such as ADCs, DACs, Codecs and DSPs. Data transmitted with I2S interface is using PCM format with high bit-depth up to 32-bit. Its first popular usage area was CD player design after it was introduced firstly at 1986 by Philips Semiconductors, but it become an industry standard shortly and now being used in almost any device which digital audio transmission or processing occurs [24].

A MEMS microphone with an I2S output includes all components exist in PDM microphones as shown in Figure 1.5. (Blog Diagram of a digital capacitive MEMS Microphone) and it additionally includes the decimation filter, therefore its output is at a standard audio sample rate which is easy to interface to and process. Microphones with I2S output can connect directly to a DSP or microcontroller with an I2S input. This saves additional components such as ADC or codec in systems and also saves the processor from the load of executing the decimation filter for both hardware and software cases.

The I2S protocol requires at least three lines including clock, word clock, and a data line. Clock line (SCK) and word clock line, typically called as word select (WS) or frame sync (FS), are driven by master device which is mostly a microcontroller. Value of FS determines channel currently active and driving data bus to send its data. Data line is mostly abbreviated with SD (Serial Data) and it includes serial bits of data coming from left and right channel repeatedly at PCM format. Typical block diagram of I2S interface is given in Figure 2.4 below.

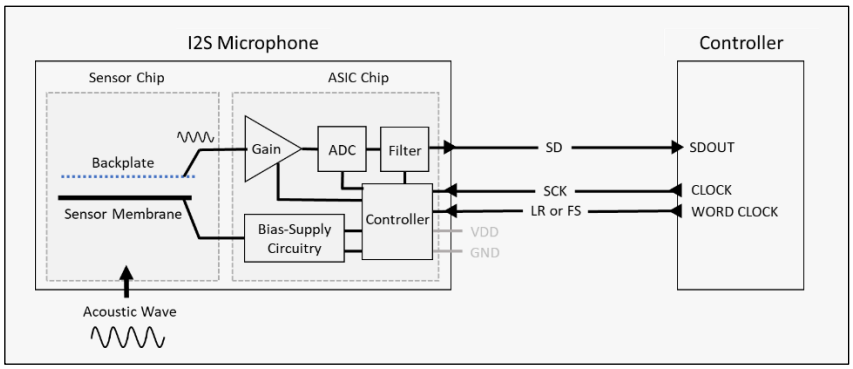

Figure 2.4. Typical block diagram of I2S interface including one MEMS microphone.

SCK line ensures synchronization between controller device (master) and microphone (slave). When there are two microphones, same SCK line feeds two of them. The second line, FS (frame sync) selects which microphone is actively speaking. Rising edge of this signal detected by the microphone which is configured to be right one using a L/R config pin. When FS is low, left microphone uses data bus whereas when FS is high, right microphone drives same bus. Typical timing diagram of I2S interface is given in Figure 2.5 below.

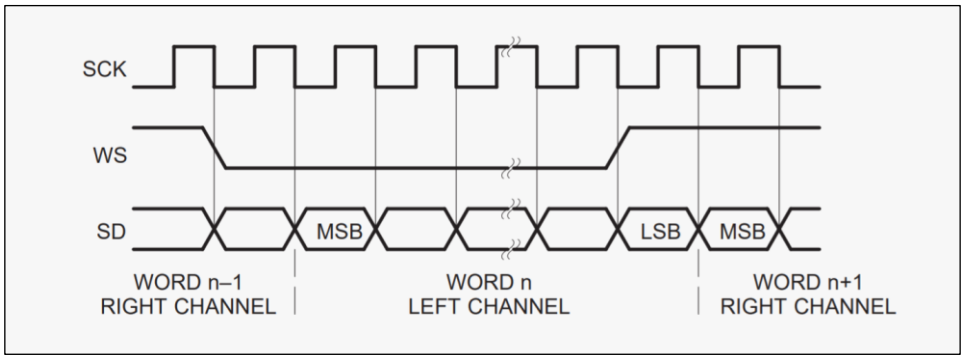

Figure 2.5. Typical block diagram of I2S interface including two channels [25].

An I2S interface can handle mono or stereo channels with a typical SCK clock rate from 512 kHz to 12.288 MHz for different sampling rates from 8 kHz to 192 kHz respectively [11]. The data word length is usually 32 bits, event actual data from microphone includes only 16 bits or 24 bits information. Unused bits are mostly filled with zero (logical low) by IC part in I2S microphones. Frame length is equal to 64-bits in most case which is twice of data word length. There are some other versions of this 3-wire configuration, which work very similar to I2S and supported by same ICs with some modifications such as position of the data word in the frame, the polarity of the clocks, or frame length. Some example versions are left-justified, right-justified, and I2S-PCM. Details of these variants of I2S not studied in this thesis since they are used very rarely.

Important advantage of I2S microphones over PDM microphones is no need for pre-decimation and conversion before processing signals digitally. This advantage becomes more important as number of microphones in mobile devices is increased whereas size of devices and ICs is reducing dramatically. Since output line is fully digital and running at lower frequencies compared with PDM format, power needed to establish communication between ICs is at lower level.

Most important limitation or drawback of I2S microphones is support for maximum of 2 channels. Increasing demand for microphone arrays containing more than 2 microphones causes need for complex multiplexing circuitry or expensive IC to merge multiple data lines of different I2S interfaces. These solutions force circuit area to be larger and therefore power consumption of overall system is increases and cost of components is increasing due to additional circuit elements placed to synchronize microphones and combine data lines.

2.4. Time Division Multiplexing (TDM)

In order to overcome maximum 2-channel limitation of I2S interface some IC producers were made some modifications on I2S microphone ICs to support easier multiplexing without external circuitry. In 2016, an important provider of MEMS sensor devices, TDK InvenSense Inc, fabricated the world's first TDM (Time Division Multiplexed) MEMS microphone supporting up to 16 microphones on one bus [26].

InvenSense's ICS-52000 low noise MEMS microphone allows connection of individual microphones directly to form a microphone array communicating with microprocessors without the need for audio converters or codecs in the system. Microphones in an array are individually able to synchronize with previous microphone in array and able to drive single data line without interference. Internal block diagram of single TDM microphone, ICS-52000 is shown in Figure 2.6. ADC, Filter and TDM serial communication IC which is also responsible for driving of frame sync for next adjacent microphone in array is included in this single device.

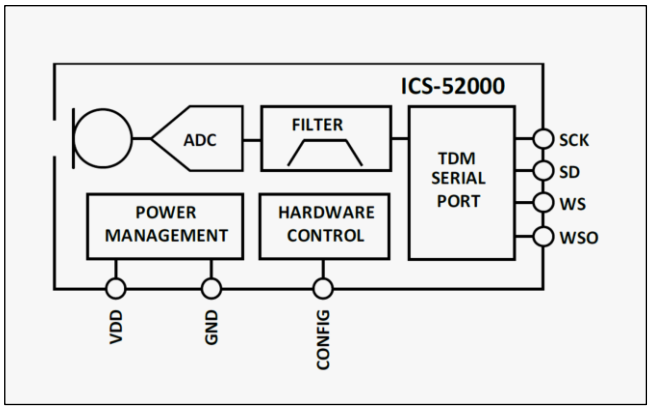

Figure 2.6. Internal block diagram of ICS-52000 TDM MEMS Microphone [27].

Contrary to I2S Microphones, there is no L/R select pin in TDM microphones. Since there is no right or left selection, function of FS input (Frame Sync) is different than I2S interface. FS input (named as WS in Figure 2.6) is used to start sampling of data from first microphone in array, and then when first microphone is finished using data bus for sending its 32-bit data, first microphone sends FS frame, same with the one it receives from master, to the second microphone in array. WSO pin shown in Figure 2.7 is connected to WS pin of next microphone is notifying next microphone when it finishes its slot operation.

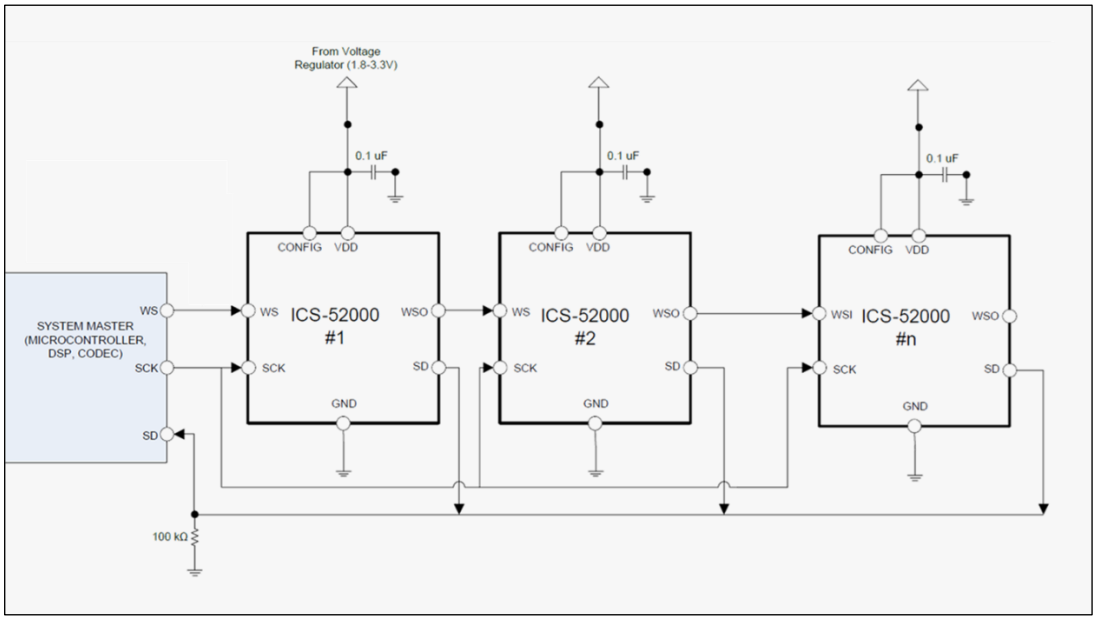

Figure 2.7. Block diagram of ICS-52000 TDM MEMS Microphone Array [27]. (Courtesy of TDK InvenSense Inc.)

This array including 3 microphones can be chained up to sixteen microphones with a single data line and each microphone drives data bus in its specific 32-bit slot in up to 512-bit frames. Pulse width or duty cycle of TDM frame sync is different than I2S's 50% duty cycle. FS pin is set to default logical low before operation started and narrow high pulse triggers first microphone to start sending its 32-bit data to bus. Length of frame is depending on N, number of microphones in array, and it can be calculated by multiplying N with 32-bit. Although it is possible to connect any number of microphones between 1 to 16 together to form an array, frame length is need to set according to N value chosen by rounding actual number of microphones to nearest power-of-two (2, 4, 8, or 16). For example, when there are 5 microphones in an TDM array actually, frame length should be selected

as 8 and last three 32-bit slots are remaining empty since data line is pulled down to logical zero when no microphone is speaking. Width of frame sync pulse can be any value smaller than frame length and single bit-wide pulse is enough to activate first microphone. The first microphone starts to drive SD data bus after rising edge of FS and continues driving for 32 clock cycle. Typical timing diagram of ICS-52000 TDM interface is shown in Figure 2.8 below.

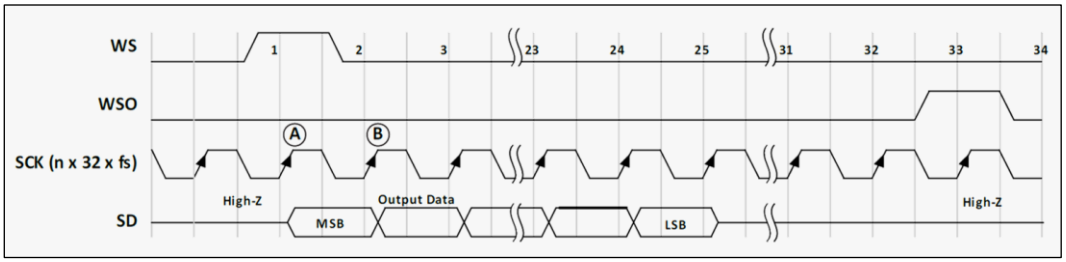

Figure 2.8. Timing diagram of ICS-52000 TDM MEMS Microphone Array [27].

First WS (Frame Sync) pulse is generated by master device which is also driving SCK clock and regenerated by first microphone to activate next microphone in array. This type of array operated by sequential communication of array elements is called as daisy-chain array. SD data bus is connected to output of all array elements directly and SD input of microprocessor.

Ensuring synchronization between SCK, FRAME SYNC (WS) and SD signals is very important. From microcontroller perspective, in order to read data in correct timing, timings of WS need to be finetuned. Since serial data is transmitted at high speed, cable length is also important factor and long paths greater than 3-5 cm is not recommended due to possible synchronization errors. If sampling frequency is 48000 Hz for an example recording application, frame sync should set at 48000 Hz as well. In this case, for 4 microphone TDM interface, SCK signal should be 4 times 32 times 48000 equals to 6.144 MHz Cable length is important for this level of communication and sync between all signals shouldn't be distorted. WS should go high at falling edge of SCK signal and data from microphones should start at rising edge (A) of clock signal. Microprocessor should read this data at rising edge of next clock cycle (B).

ICS-52000 microphones have bit-depth of 24 whereas each slot consist of 32-bits little endian signed integer data values. Data format at each 32-bit slot consist of two 16-bit parts representing MSB (most significant bit) and LSB (least significant bit) parts respectively. Left 16 bits of microphone output fills MSB part and remaining 8 bits fills first 8 bits of LSB part. Remaining bits

in LSB part is filled with zeros since no data is pushed from microphones at this interval. Resulting data format available at single slot is big endian byte order 32-bit signed integer as shown in below Figure 2.9.

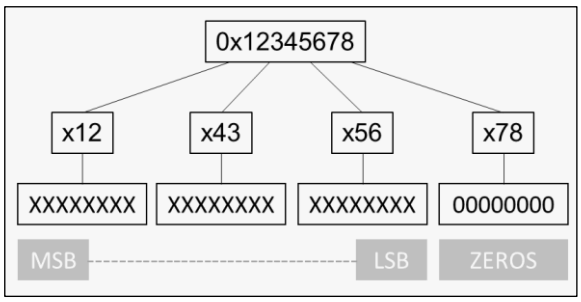

Figure 2.9. Output data format of ICS-52000 TDM MEMS Microphone.

Since ICS-52000 fills last byte with zeros by itself, microprocessor can directly accept and record incoming 32-bit data without shifting but, since original bit-depth is 24-bit and there will be no data loss when division of integer to 128 (8 bit shift) occurs, 8 bit shift before recording samples can be done at software level by driver responsible for data parsing and calculations.

TDM interface has no standard yet whereas the I2S has Philips standard. Therefore, different IC manufacturers adding some modifications to the original TDM interface according to needs of their applications. These variants usually derived from different clock polarities, tri-stating and driving method of unused channels. System designers should be careful to verify that output format of an TDM array element is matches with data format that is expected by microprocessor or other processing units. Since it is a relatively new technology, support for TDM from microcontroller perspective is very limited. Hardware and microprocessors supporting TDM interface will be discussed in next section.

2.5. Supported Interfaces by Platforms and Microcontrollers

In this section, different microcontrollers and SoCs (system-on-a-chip) will be evaluated from supported audio interface perspective. PDM interface is most common interface supported by almost all legacy and modern microprocessors whereas TDM is only supported by a few platforms currently. Even hardware has capability to drive TDM communication practically, low-level peripherals and

drivers should allow developers to configure and run CLK and WS signals properly. Table 2.1 demonstrates summary of supported interfaces by different micro controllers and platforms.

Table 2.1. PDM, I2S and TDM support of platforms and microcontrollers [28-34].

Manufacturer	**Microcontroller**	**On Platform**	**PDM Interface**	**I2S Interface**	**TDM Interface**
Broadcom	BCM2837	Raspberry Pi 3b+	Supported	Supported	-
Microchip Technology	SAMD21	Arduino MKR Zero	Supported	Supported	Supported
Microchip Technology	SAMD21G18	Arduino Zero	Supported	Supported	Supported
Microchip Technology	ATmega328P	Arduino Uno	-	-	-
XMOS	xCORE-200 XU216 multicore microcontroller	xCORE Microphone Array board	Supported	-	-
XMOS	VocalFusion XVF3000 mono-AEC voice processor	-	Supported	Supported	-
Freescale Semiconductor, Inc.	MK20DX256VLH7	Teensy 3.2	Supported	Supported	Supported
Microchip Technology	SAMD21/DA1 Inter-IC Sound Controller	-	Supported	Supported	Supported

There are some additional codec ICs available in the market to fulfill support need for these interfaces in case of controller itself is not containing required interface. CS4244 4-in/4-out Audio Codec by Cirrus Logic, Inc. is an example codec containing analog-to-digital convertors, multi-bit digital-to-analog Delta Sigma converters and TDM interface. It is also compatible with differential inputs including PDM and other output formats [35].

2.6. A workaround: Adding TDM support to I2S interface

During component selection and device design phase, there are many different constraints including cost of device, internet connectivity, processing power, input/output ports, and required firmware etc. Although some platform options are not supporting TDM interface, it is not easy to eliminate advantageous options offering high computing power, low-cost, dual-band Wi-Fi support etc. Since I2S and TDM interface are very similar in terms of slot size, frame sync and clock signal, it can be tried to use I2S interface of a microcontroller with some additional components to make necessary modifications between I2S and TDM.

BCM2837 SOC of Broadcom has almost all of required features but it is supporting only I2S not TDM interface. It can be used to design such a workaround thanks to its high computing power and flexible I/O capability. BCM2837 offers advanced configuration options for I2S interface including frame length, channel width, channel position and frame sync pulse length etc. as shown in Figure 2.9.

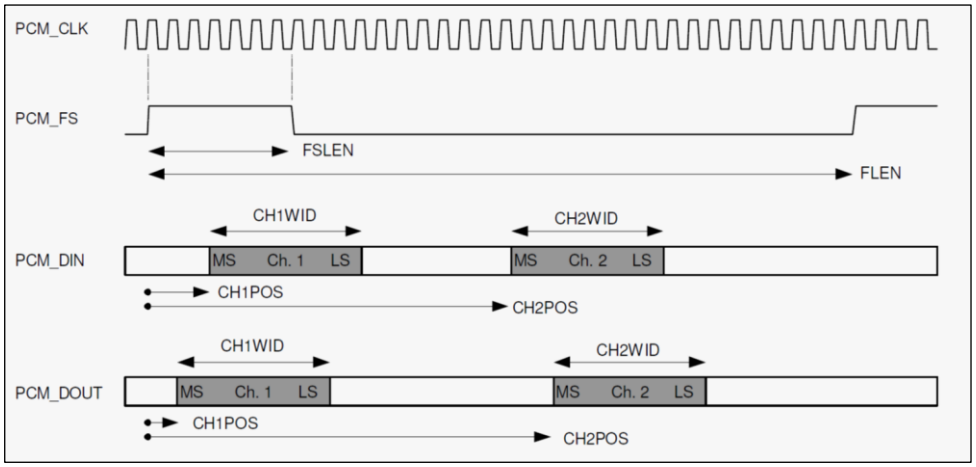

Figure 2.10. Timing diagram and configurable parameters in I2S interface of BCM 2837 [28].

Most difficult limitation on modifying this interface is number of channels available. I2S interface supports maximum two channels. Handling generation of FS signal which is suitable for TDM microphones is possible by utilizing a 4-bit binary counter. Original IS2 interface and additional circuitry responsible for FS conversion is shown in Figure 2.9. Noting that TDM data flow starts on rising edge of FS, whereas I2S dataflow starts on falling edge of FS, counter is triggered on negative edge of I2S_FS and counts for 2 edge to generate next TDM_FS_4MIC pulse. Therefore, 4 microphones will use bus subsequently and microprocessor waiting for 2 channel data will get all 4 slots as two subsequent stereo frames. Basic operation on software can correct this format to original TDM format which is 4-channel 32-bit PCM. Same methods can be also applied to design 8 or 16 microphone TDM interfaces, using increasing counter steps as shown in Figure 2.9.

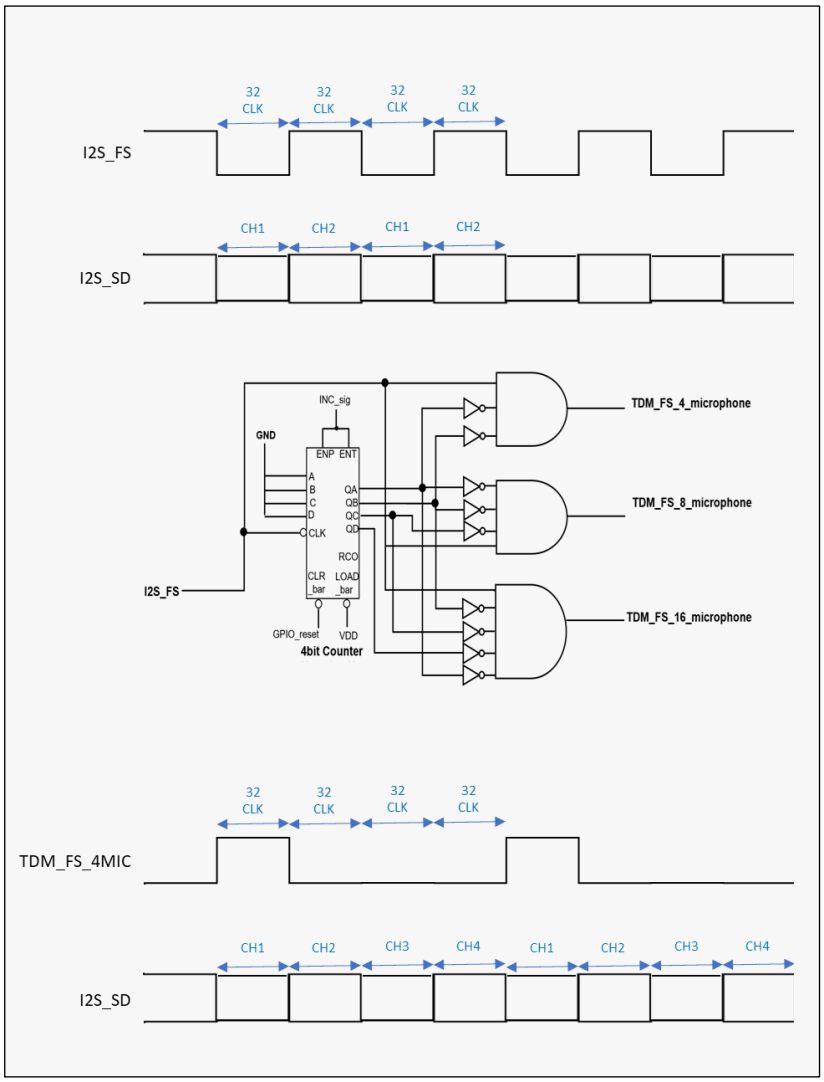

Figure 2.11. Block and Timing diagram of I2S to TDM conversion for 4-Mic Array.

3. DESIGN AND EXPERIMENTS

In the first phase of thesis study, the main focus was to learn and test MEMS microphone alternatives deeply before designing a completed microphone array device consisting at least 4 microphones. After reviewing related literature and products from different manufacturers, short list of alternatives is ordered for further practical testing on them. As detailly mentioned in previous chapter, TDM support is relatively new technology but in brings many advantages including no need for additional circuitry while allowing up to 16 microphones. Therefore, the final array device is aimed to use TDM interface and other alternatives in addition to TDM microphones are detailly reviewed.

3.1. Evaluation of Microphone Alternatives

The first microphone selected to be tested was ICS-43432 digital MEMS microphone from InvenSense Inc. and using I2S interface. It is a bottom port microphone as shown in Figure 1.2 consisting of an acoustic sensor, an analog-to-digital converter, decimation and anti-aliasing filters, power management blocks. Thanks to advantages of I2S interface mentioned in previous chapter, it can be directly connected to microcontrollers without intermediate codec need in device. Raspberry Pi Zero and Raspberry Pi 3b+ are used to test ICS-43432 microphones.

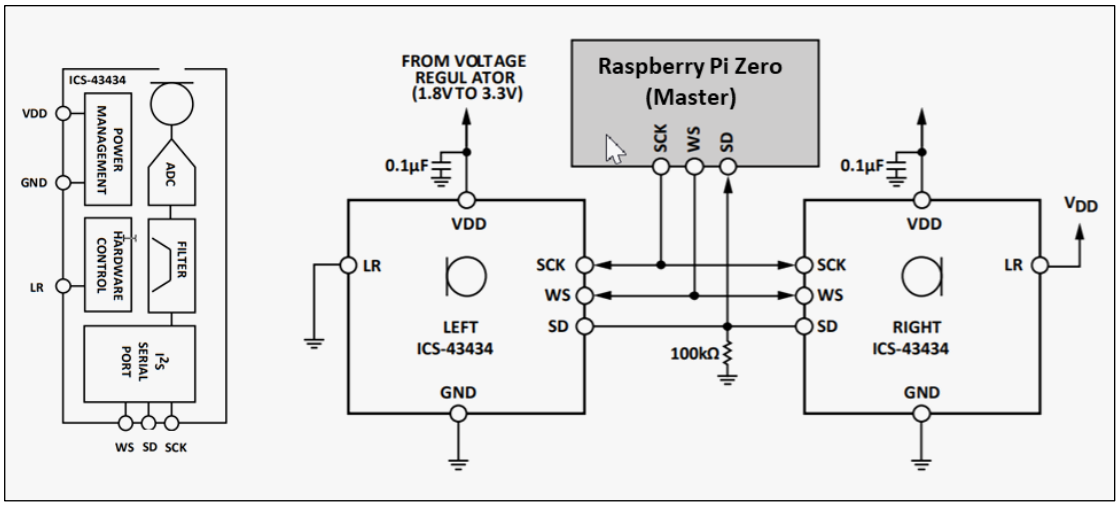

Figure 3.1. Block diagram of I2S interface between ICS43432 and Raspberry Pi Zero [36]

Since there is built-in support for I2S interface on Raspberry Pi firmware and on underlying SoC (BCM2837), recording of two channels (stereo) was done using standard audio libraries exist in ALSA (Advanced Linux Sound Architecture). Starting phase of 1000 Hz sinusoidal tone was captured as shown in Figure 3.2.

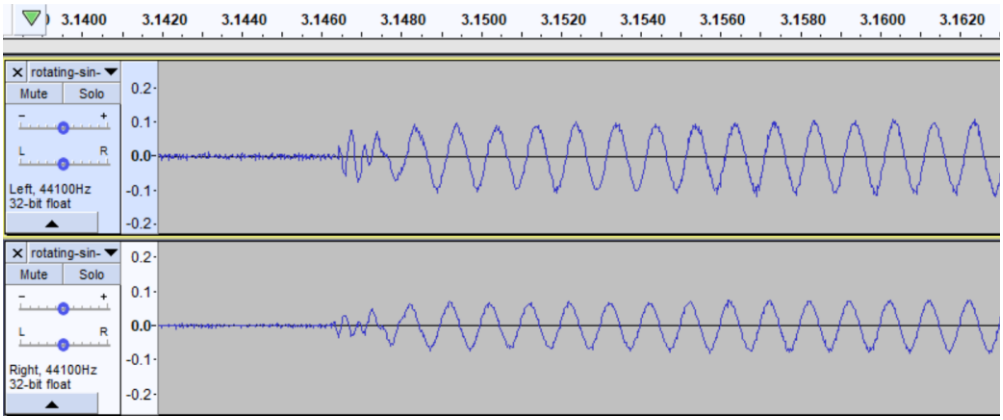

Figure 3.2. Sample stereo recording of 1000 Hz Sin Wave at 44.1Hz using ICS43432.

According to datasheet of ICS-43432, it has a high signal-to-noise ratio of 65 dBA and sensitivity of -26 dB [36], which is suitable for high-performance microphone arrays without necessity of calibration but, I2S protocol allows maximum 2 microphones to communicate with a master microcontroller. To create a 4-mic array with I2S microphones, additional codecs, or circuit components are required. MAX9880A Audio Codec by Maxim Integrated is an example codec supporting dual I2S inputs and TDM interface together with additional digital filtering capabilities [37]. During this study, number of components required to create array was aimed to be minimized and other commercial products are continued to be reviewed.

One of most common microphone array shields used by researchers and hobby users is ReSpeaker 4-Mic Array by Seeed Studio. It consists 4 MEMS Analog microphones and an ADC (Analog to Digital) convertor (AC108) with 4 input support. It basically transmits 4-channel digital sound data to Raspberry Pi via I2S interface by dividing it to 2 pieces of stereo recordings in software codec. Analog MEMS microphone inside Respeaker, SPU0414HR5H-SB is product of Knowles Electronics which is the largest MEMS microphone manufacturer supplying more than half of microphone market worldwide and offering high-performance microphones for smart phone industry including Samsung and Apple [38].

Last but not least, ICS-52000 microphone by InvenSense using TDM interface was also reviewed and compared with other alternatives as shown in Table 3.1. Although ICS-52000 offers good sensitivity and SNR characteristics, avg. current value is higher than other alternatives. This is result of daisy chain operation of microphones as explained in section 2.4. Since each microphone is responsible for driving WS for next microphone and includes internal circuitry to enable this function this current value is at acceptable level.

Table 3.1. Comparison of reviewed and tested MEMS microphones [27][36][39][40].

Specification	**ICS-43432 by InvenSense**	**ICS-52000 by InvenSense**	**SPU0414HR5H-SB by Knowles**	**SPH0645LM4H-B by Knowles**
Output	Digital I2S	Digital TDM	Analog	Digital I2S
SNR	64 dBA	65 dBA	59 dBA	65 dBA
Sensitivity	-26 dB FS	-26 dB FS	-24 dB FS	-26 dB FS
Avg. Current	230 - 490 µA	1.1 mA	155 µA	600 µA

After review of microphone alternatives, ICS-52000 digital microphone was selected to continue for final PCB and device design thanks to the support for up-to 16 microphone and less peripheral need.

3.2. Tests with Evaluation Boards

In order to run ICS-52000 microphone with a microprocessor before producing a PCB, NW-AUD-ICS52000 evaluation board from NotWired.CO was ordered. This evaluation board provides an easy to use, low-cost, linear 4 microphone array configuration [41]. Female and male connectors on both side of PCB allow board to be connected with another board as daisy chain tiles. When up to four pieces of the board is connected end-to-end, linear and uniformly distributed 16-microphones array can be formed easily. Schematics and real photos of this board are shown in Figure 3.3.

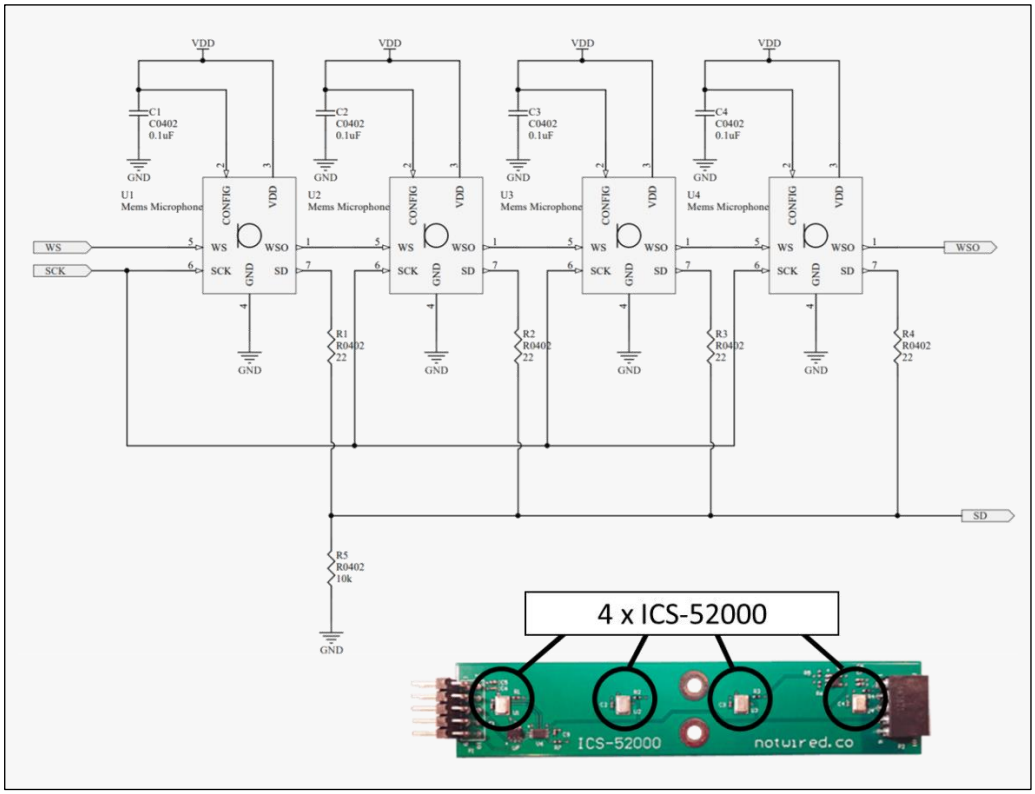

Figure 3.3. Schematics (important parts) and PCB of AUD-ICS52000 evaluation board [41].

After this evaluation board was procured, microprocessor alternatives are reviewed detailly. Raspberry Pi 3 was preferred as first option because of high performance features which will be required for real-time signal processing and wireless communication needs on the aimed final device. Raspberry Pi operates on BCM2837 SoC by Broadcom and this chip supports only 2-channel I2S interface by default. In order to operate TDM microphone array using this chip, some modifications on software level are done including writing an ALSA SoC Digital Audio Interface driver and CLK and FS signals are successfully generated according to TDM interface requirements. First change needed to enable TDM communication was inversion of FRAME_SYNC signal since I2S starts working on falling edge of FS whereas TDM starts on rising edge of FS as it mentioned detailly section 2.4. This change was done using MODE_A Register of PDM module exist in BCM22837. In order to change direction of FS, FSI (Frame Sync Inverse) field of MODE_A Register was changed to zero instead of one. Another change required related frame sync signal was length of FS frame. It was also configured using FLEN (Frame Length) field of MODE_A Register. It was set to 128-1=127 bits to support 4 continuous 32-bit slots in it. Last change done using registers was length

of FS pulse. It was set to 32-bit using FSLEN (Frame Length) field of MODE_A Register even it was not very important for TDM interface since it waits only for rising edge. After making all these modifications and changes in Linux sound card to support 4 channels, CLK and FS signals started to work as desired and data flow was started on SD line. When data coming from microphones was recorded and analyzed but nothing more than a noisy signal was observed. Figure 3.4 shows oscilloscope picture of working configuration supplying 4 x 32-bits x 44.1 kHz = 5.64 MHz for 4 channels.

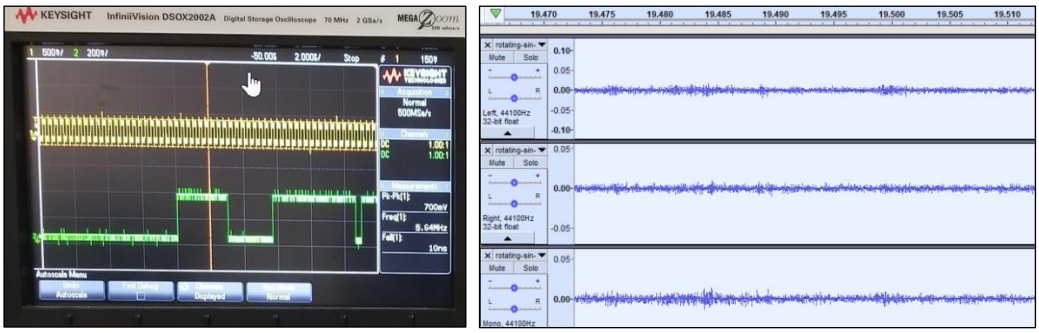

Figure 3.4. CLK (Yellow) and SD line (Green) on left and resulting noise for sinusoidal 1000 Hz tone recording using AUD-ICS52000 evaluation board.

In order to make it work, tens of variations on configurations are tried and controlled but result doesn't change at all. Suspected from a defect on MEMS microphones inside board and different boards are tested with same setup but result not changed. To ensure microphones are working properly, another microprocessor alternative, Teensy 3.2 was procured because it has native support for TDM interface.

The Teensy is an Arduino based complete microcontroller development system, in a very small footprint, capable of executing many types of projects including sound interface applications. Programming of Teensy boards is done via the micro USB port similar to Arduino and Teensy 3.2 features a 32-bit ARM processor capable of both I2S and TDM.

Verified code examples from Teensy community reviewed and exact same configuration as is in tutorials was setup but result was obtained same with test hold with Raspberry Pi Zero and 3b+. Suspecting from eval board internal circuitry, individual microphones inside board was separated and interface was tested with single microphones. Same noise signals obtained for single

microphones too. Coupling effects, delays because of long cables and all other potential causes of error investigated deeply and ordered individual ICS-52000 microphones from manufacturer to make all tests one-by-one without using evaluation board. After procurement of bare microphones, a microphone was soldered with very thin copper lines and connected to Teensy 3.2. Finally, proper sound pulses are observed in both oscilloscope and recorded data as shown in Figure 3.5.

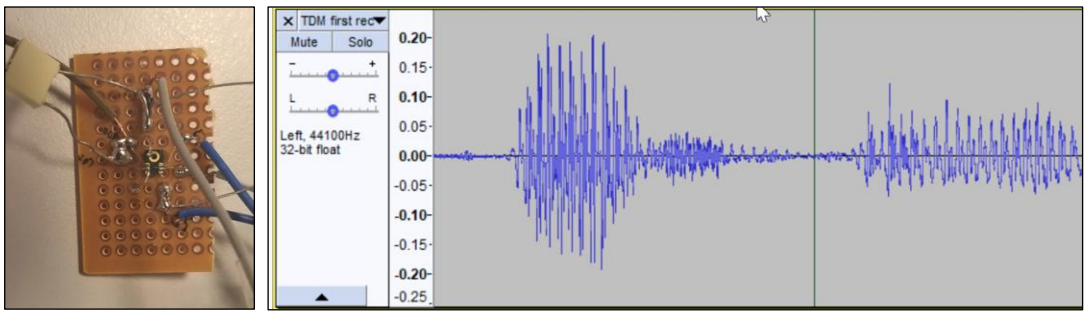

Figure 3.5. Soldering of bare single ICS-52000 microphone (on the left) and successfully recorded human voice pulses using Teensy 3.2 (on the right).

After getting correct and proper sound signals from this single microphone setup, two ICS-52000 microphones from both AUD-ICS52000 evaluation board and new ordered ones are compared to reach final evidence about failures previously faced. It was realized that series of these two group was different, and it was understood that all microphones in all AUD-ICS52000 evaluation boards was broken or damaged. Figure 3.6 shows very small difference between serial numbers of working and broken microphones.

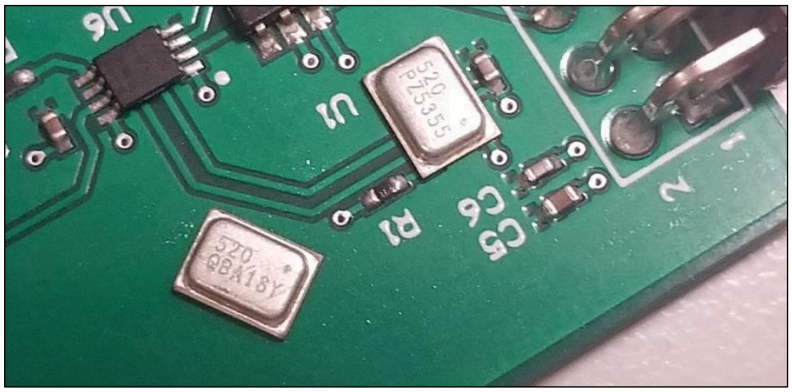

Figure 3.6. New ordered ICS-52000 microphone (520QBA18Y) on the left and ICS-52000 on AUD-ICS52000 evaluation board (520PZ5355) on the right.

After single microphone is working accurately, 4-microphones array consisting new and working ICS-52000 microphones was prototyped using a square shaped epoxy board as shown in Figure 3.7.

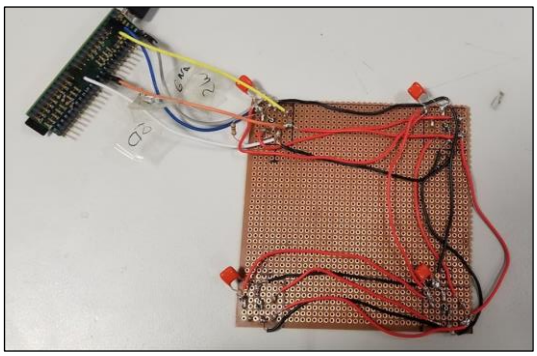

Figure 3.7. Prototype array build on top of epoxy board consisting new ICS-52000 microphones connected to Teensy 3.2 board.

It was managed to retrieve real-time data from the microphones through serial monitor of Teensy 3.2. During tests with this prototype, small distortions in signals are started to be observed at high sampling rates, especially above 16kHz. To investigate and solve this issue, clock and data lines are replaced with coaxial cables as first step to decrease possible distortions. Second modification made was changing clock path to form a spider distribution network instead of rotating path and adding additional resistors at each end of clock lines to match them with 50-ohm transmission lines as suggested by Johnson and Graham, 1993 [42]. After this steps, high frequency without distortion limit was increased to about 32 kHz. Further investigation is required for perfect signal forms.

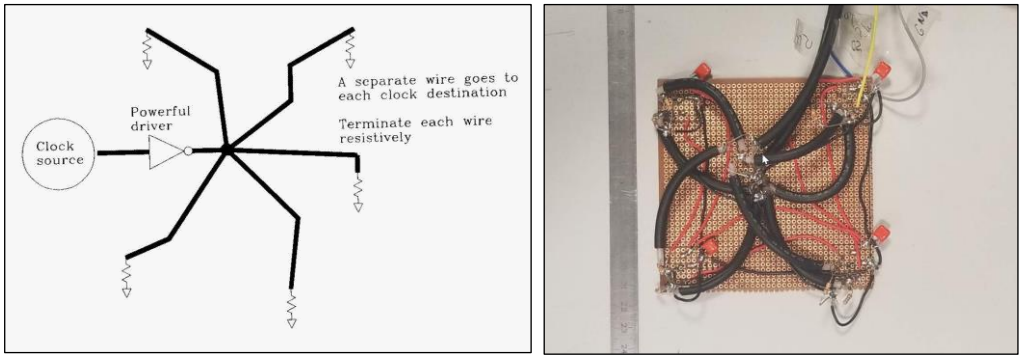

Figure 3.8. Spider Distribution Network [42] and implemented prototype.

3.3. PCB Design and Production

Aimed microphone array device during this study needs high computing power to process real-time data and to made required communication with external systems. Teensy board has some drawbacks in terms of lack of application layer allowing python codes to run simultaneously and wireless connectivity. Therefore, final device PCB was designed to fit and easily plug with Raspberry Pi Zero or Raspberry Pi 3b+. Number of microphones required was set at 4 and shape of array was defined as square with edges 75 millimeter each. Land pattern layout of ICS-52000 was obtained from its datasheet as shown in Figure 3.9.

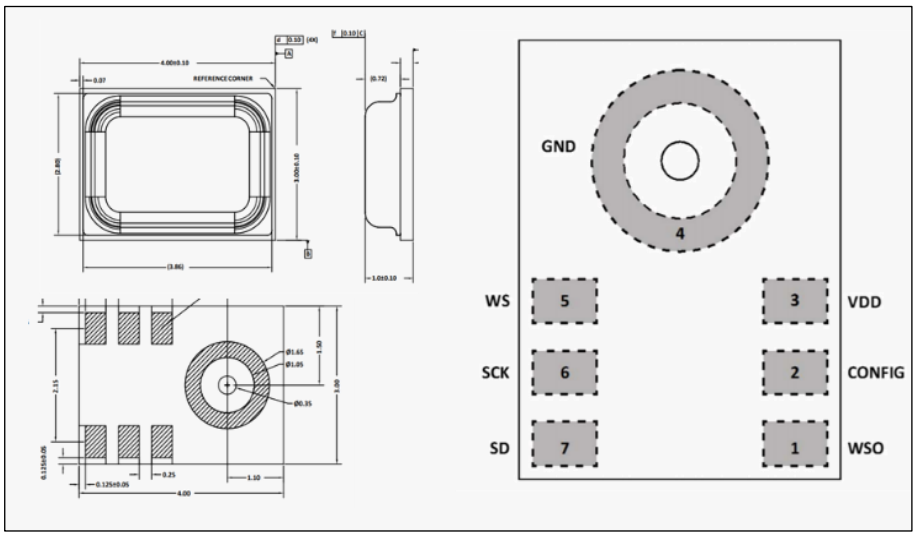

Figure 3.9. 3D Dimensions and PCB Land Pattern Layout of ICS-52000 [27].

3.3.1. Schematics and Layout Design

Overall circuit schematics was drawn on Eagle according to recommendations mentioned in datasheet of ICS-52000 before working on PCB layout. All passive components including resistors and capacitors and active components such as buffers are added. Additionally, ESD (Electro-static Discharge) Protection module was placed to protect microphone data from potential electro-static charges. Final version of device schematics is shown in Figure 3.10.

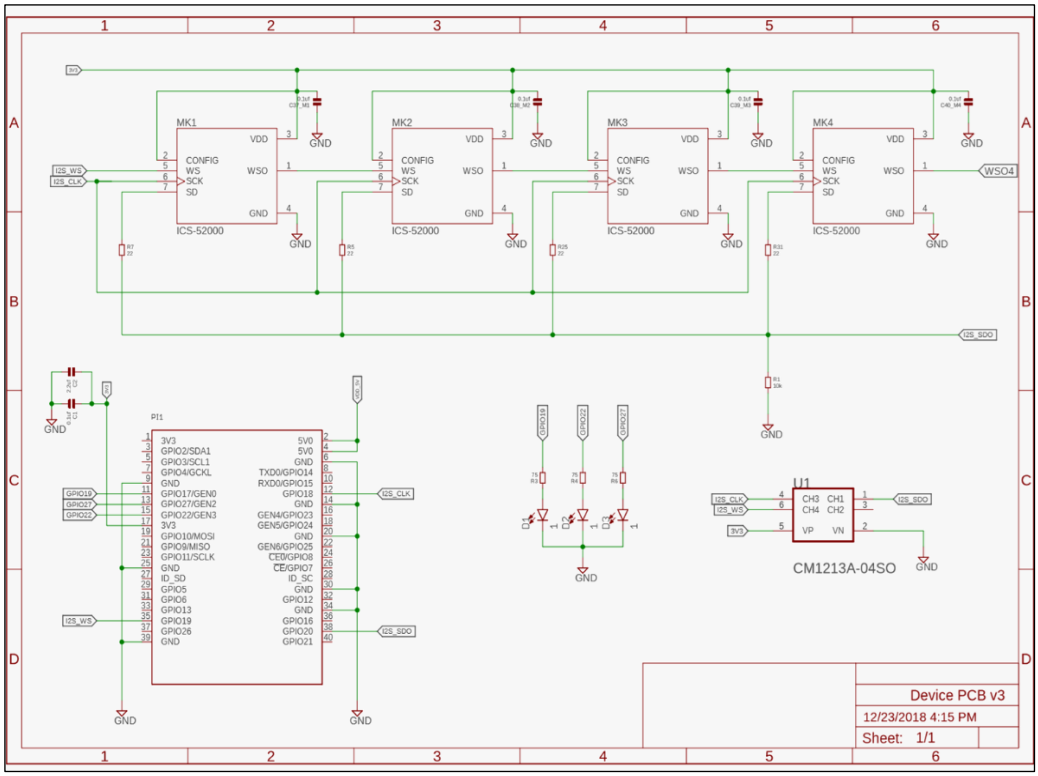

Figure 3.10. Schematics of Designed Microphone Array Device.

In this schematics, important components are 4 pieces of MEMS Microphone (ICS-52000), 3 pieces of status LED, an Electro-Static Discharge protection module, 40 Pin Header (for Raspberry Pi connection) and many numbers of resistors and capacitors. After all components are placed and connected properly, ERC (Electrical Rule Check) feature in Autodesk EAGLE was run to check possible connectivity errors.

As next step, switched to PCB Layout section of Eagle CAD, and all components exist in schematics are precisely located on defined shape of PCB. In order to maximize array size while not enlarging overall device size, PCB was shaped in a way that headers and USB ports of Raspberry Pi 3b+ fills cavities of PCB to form compact square shaped device.

Designed PCB layout and 3D placement of PCB into the final device box is given in Figure 3.11 below. Four MEMS microphones are placed extensions on each corner and three status LEDs are placed in center of the board. After PCB design DRC (Design Rule Check) feature in Autodesk EAGLE was run to check possible connectivity errors and all of possible errors are fixed.

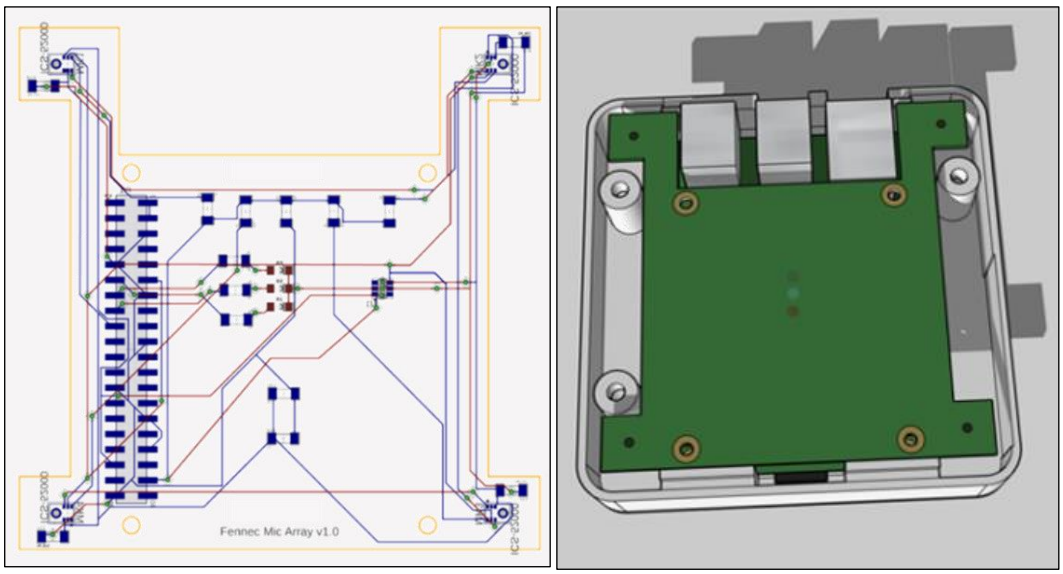

Figure 3.11. 2-Layer PCB Layout and designed PCB placement in 3D device box.

3.3.2. PCB Production and Assembly

After designed PCB and schematics are checked for possible layout errors, PCB layout is exported as Gerber format which is an open 2D binary vector image file format used by printed circuit board industry to describe the printed circuit board layers including copper, solder mask, legend, and labels etc. For fast and cost-effective production, Gerber files were submitted to Seeed Fusion PCB printing service and printed PCB was acquired very rapidly. Assembly of all components to the PCB was done by hand for the prototype device.

Acquired PCB is checked for possible short-circuits before soldering components. Then all passive and active components are soldered properly to the PCB and circuit is powered to make initial tests. Assembled board after soldering is shown in Figure 3.12.

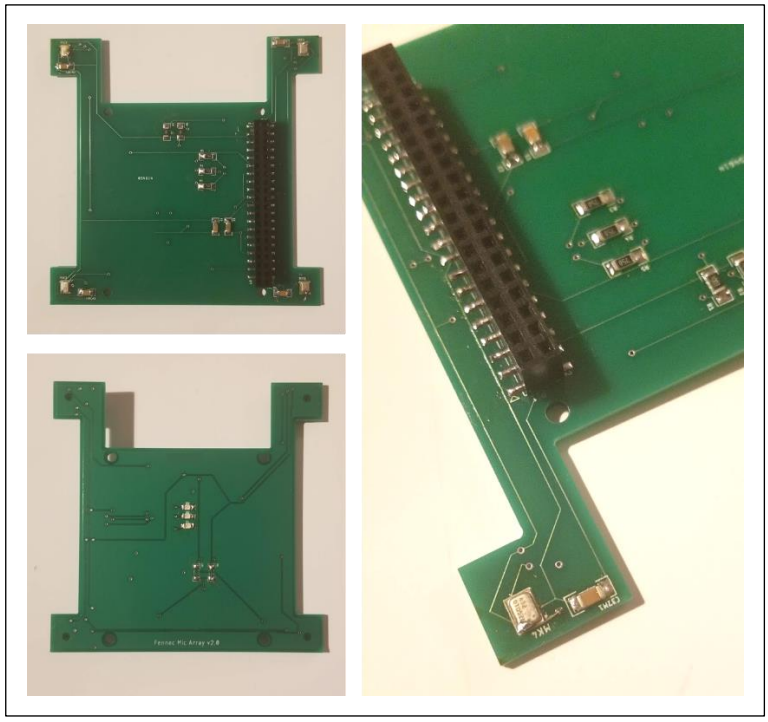

Figure 3.12. Assembled PCB of Designed 4-Mic Array Device.
(Bottom, Top and Perspective Views)

3.4. Sound Source Localization

Different sound source localization methods were studied in Section 1.3 detailly. In this section of thesis study, MEMS microphone array which was designed and produced as mentioned in previous sections will be used in a sound source localization application and performance of localization method using this array will be examined.

4-mic square array designed has both end-fire and broadside characteristics depending on potential placement options of sound source. This information is unknown in most real-life environments, but there are some environmental constraints that are tied to desired scenario or application. As an example, for a voice localization and recording application required on a teller desk as shown in Figure 3.12, height of teller desk and position of teller relative to microphone array position are two important parameters which are mostly known. While exact height or mouth position of teller and customer is unknown, a range can be defined that will cover most cases.

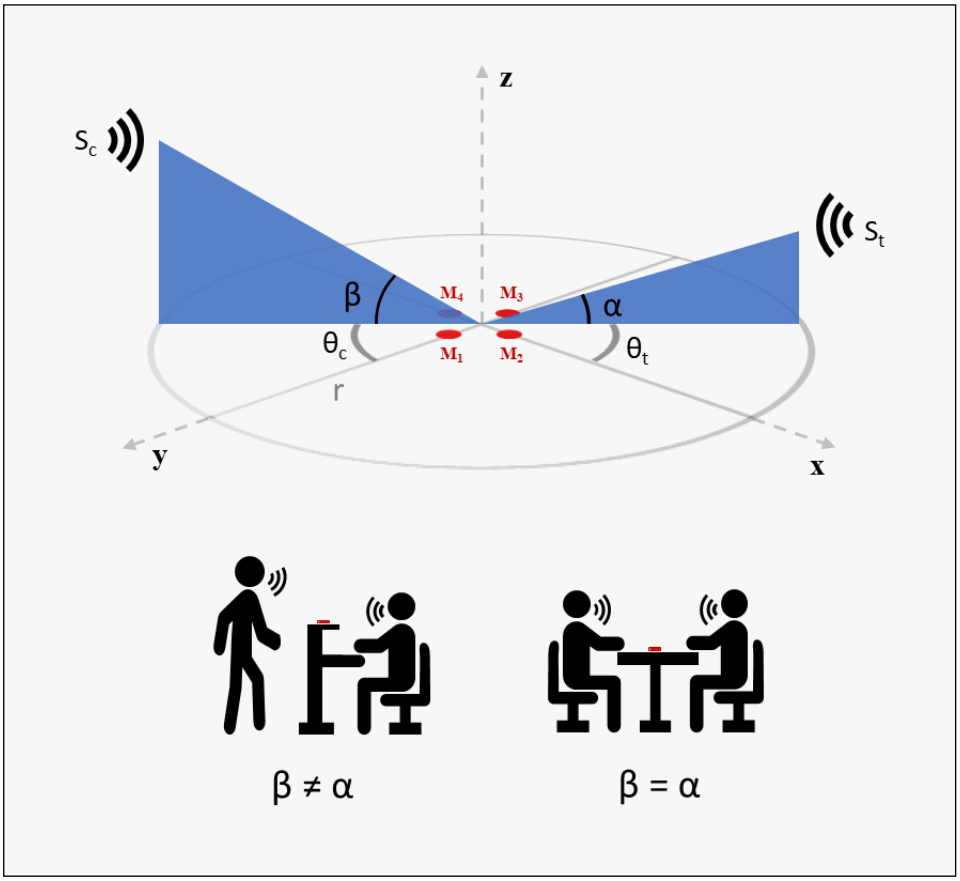

Figure 3.13. Teller Desk with Microphone Array Scenario Representation.

In this scenario, teller and customer are two potential sources of sound, S_t and S_c respectively. Microphone array (M1, M2, M3, M4) is located on top of desk is located at center of circle which is intersecting both teller and customer. As it explained in different array geometries section 1.3, M1-M2 and M3-M4 microphone pairs are forming end-fire array geometry whereas M1-M4 and M2-M3 pairs are forming broadside array geometry.

Another constraint of this scenario is acoustic characteristics of target sound sources which are desired to be located. Fundamental frequency of the voice of a male adult is ranging from 85 to 180 Hz and the voice of a female adult is ranging from 165 to 255 Hz [43], but harmonic series present in recording usually causes up to 3.4 kHz actual frequency range. Voice spectra of males, females and children (1/3 octave) depending on efforts, studied by Brixen, Eddy B., is shown in Figure 3.13 [44].

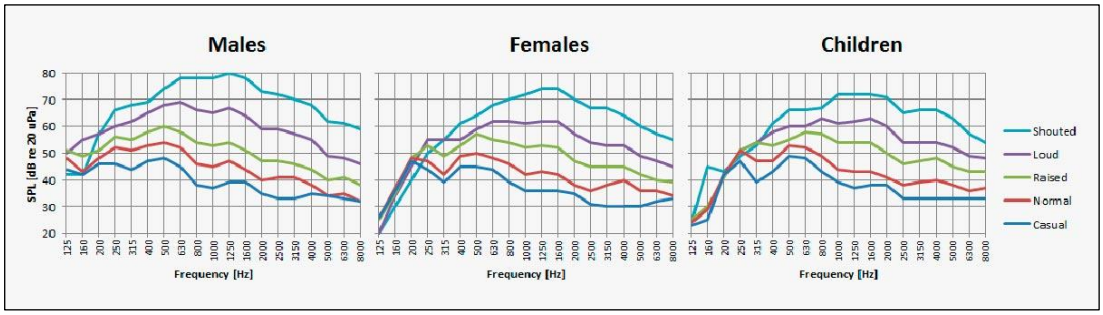

Figure 3.14. Voice spectra of males, females and children (1/3 octave) by effort [44].

In the experiment setup, in order to asses success of method used quantitively, reference sound source was selected as recording of a keyword pronounced by adult male. Sound source was rotated at circle defined as $x^2 + y^2 = r^2$ and z = h, where r = 50 cm and height of sound source relative to microphone array, h = 20 cm.

During data from all 4 microphones were recording, signal all possible pairs were analyzed real-time using time delay of arrival method and compared with other pairs to calculate direction of arrival angle of dominant sound source in each time interval.

3.4.1. Time Delay of Arrival using Cross-Correlation

As a first step, sound source repeating record of a keyword was placed at 0° point of circle as shown in Figure 3.14 and started rotating clockwise with 45° steps. Full rotation was completed, and 4-channel data was collected successfully as shown in Figure 3.15.

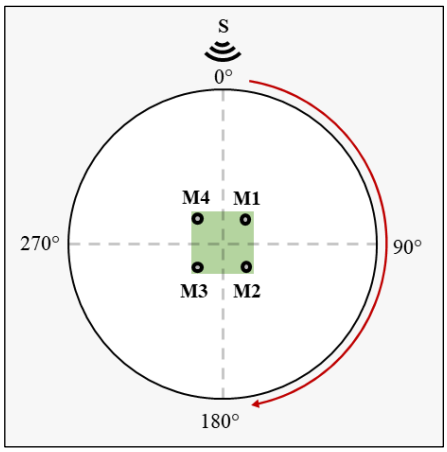

Figure 3.15. Direction of arrival experiment, top view.

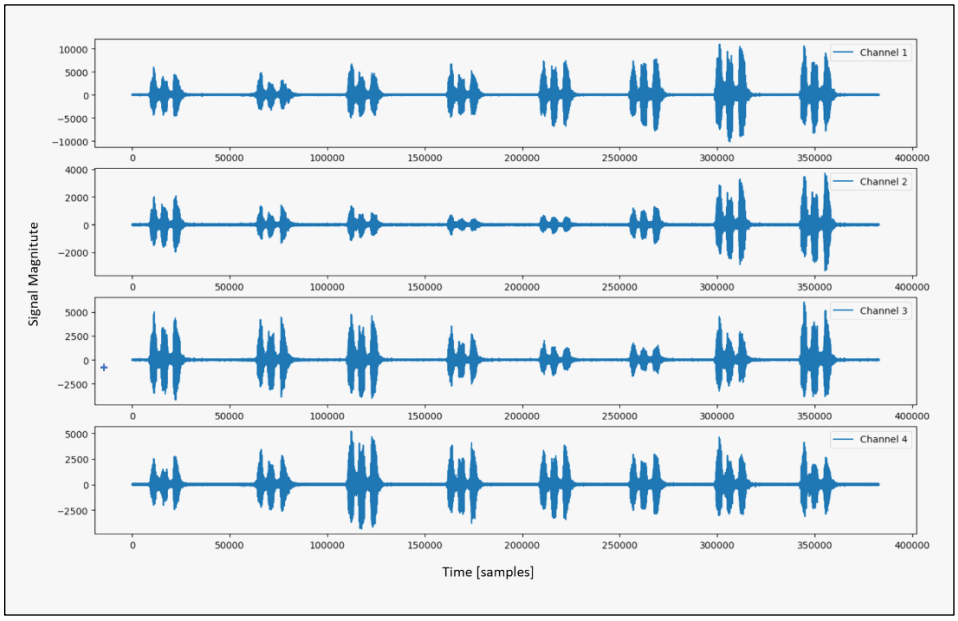

Figure 3.16. Recorded data at 44.1 kHz for direction of arrival analysis.

All channels were normalized to ±1 amplitude range in order to minimize power differences due to the small obstacles affecting individual channels. After normalization, voice activity was detected using simple threshold method as shown in Figure 3.16. Blue regions in the figure shows part of signals passing above the threshold value of 0.2.

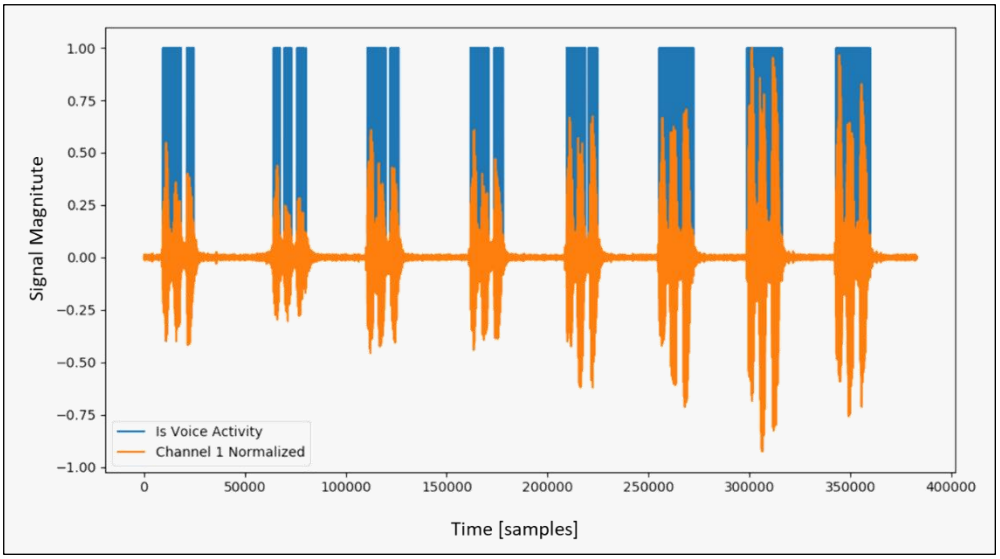

Figure 3.17. Voice activity detection.

As a next step, channel pairs were correlated using GCC-PHAT Cross-Correlation method. Microphones M1 and M3 were formed first pair and remaining M2 and M4 was formed second microphone pair. Cross-Correlation C_d of two signals S_1 and S_3 were calculated as following equation where k is representing all possible values of time delay τ between microphones.

$$\text{Delay}_{max} = \frac{Distance\ Between\ Microphone\ Pair\ [m]}{Speed\ of\ sound\ in\ Air\ [m/s]} = 0.329\ milliseconds \quad (3.1)$$

Using Delay_{max} value and sampling frequency, which is 44.1 kHz, possible k values should be less than or equal to 0.329 x 44100 = 14 and greater than or equal to -14. Then using moving 200-sample window which includes voice activity, correlation array was obtained for each window as shown in Figure 3.16.

$$C_d[k] = \sum_{n=0}^{N+200} S_1[n].S_3[n+k] \quad where -14 > k > 14 \quad (3.2)$$

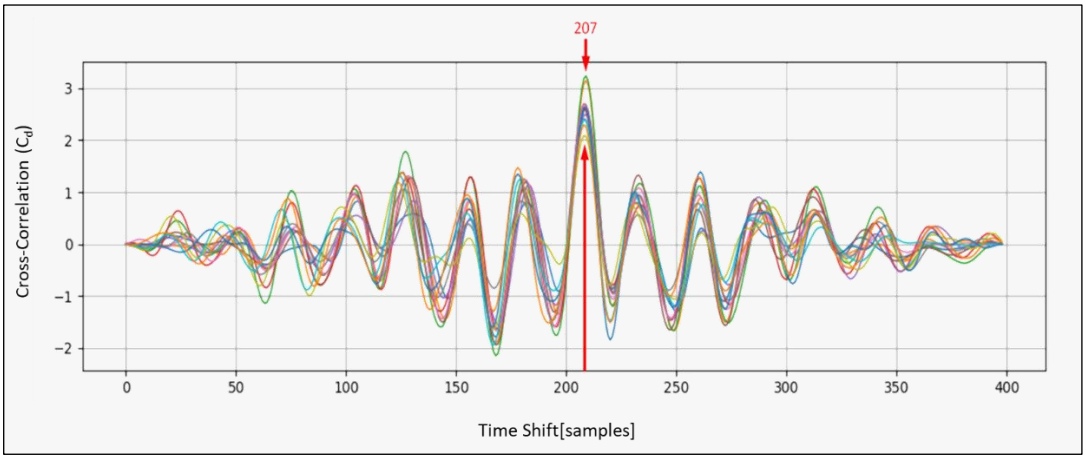

Figure 3.18. Cross-correlation arrays for different 200-sample windows for same DoA.

As it seen from Figure 3.15, highest values of correlation were observed on 207th element of the correlation array. Since correlated arrays have 200 sample elements, this means 7 sample shift occurs between selected channels (S1 and S3) due to the time delay of arrival. Positive sign of this shift means that microphone 1 is closer to the sound source than microphone 3.

3.4.2. Direction of Arrival Estimation

As a next step, time delay between other pairs was also calculated using same approach and utilizing $\tau = argmax(C_d[k]) \,/\, f_s$ formula. As a result, delay 1-3 (between microphones M1 and M3) and delay 2-4 (between microphones M2 and M4) were obtained as shown in Figure 3.16. Single signal passing through x-axis in background is representation of channel 1 in order to compare horizontal location of points in the figure.

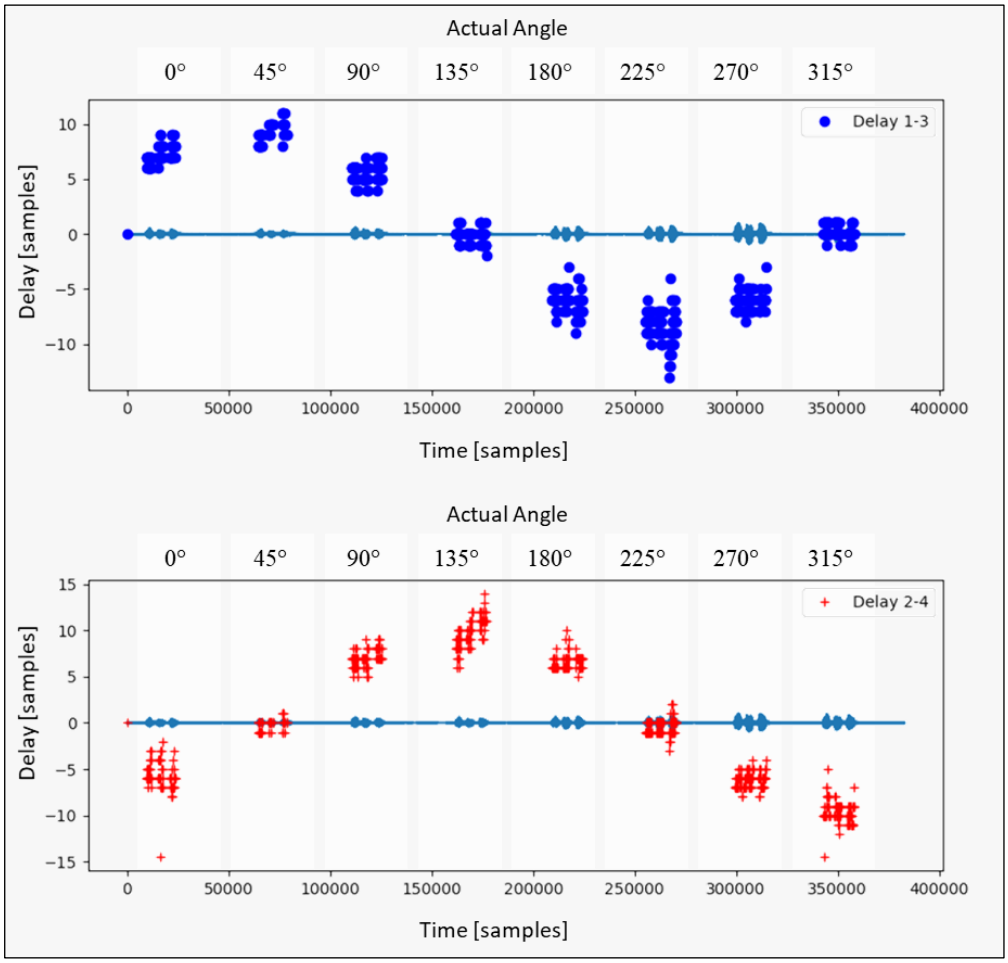

Figure 3.19. Time delay calculation between microphone pairs.

Then using geometry of array and following equations 3.3 and 3.4 where delay_max is equal to maximum delay duration between far most microphones, direction of arrival angle for each sampling window is calculated.

$$\theta_{13} = \text{asin}\left(\frac{\text{delay_13}}{\text{delay_max}}\right) * \frac{180}{\text{pi}} \quad \textbf{and} \quad \theta_{24} = \text{asin}\left(\frac{\text{delay_24}}{\text{delay_max}}\right) * \frac{180}{\text{pi}} \tag{3.3}$$

$$\begin{aligned} \text{DoA } [^\circ] = 120^\circ - (& \\ & \textbf{if } |\theta_{13}| < |\theta_{24}| \textbf{ then} \\ & \quad \textbf{if } \theta_{24} > 0 \textbf{ then } \theta_{13} \bmod (360) \textbf{ else } 180 - \theta_{13} \\ & \textbf{else} \\ & \quad \textbf{if } \theta_{13} < 0 \textbf{ then } \theta_{24} \bmod (360) \textbf{ else } 180 - \theta_{24} \\ &) \end{aligned} \tag{3.4}$$

Resulting DoA (direction of arrival) angle and related actual positions are shown in Figure 3.17. There are two separate regions in 5th group of signals which is belong to 180° arrival angle. Possible causes of this imperfection are non-ideal environment and not removal of points by the algorithm.

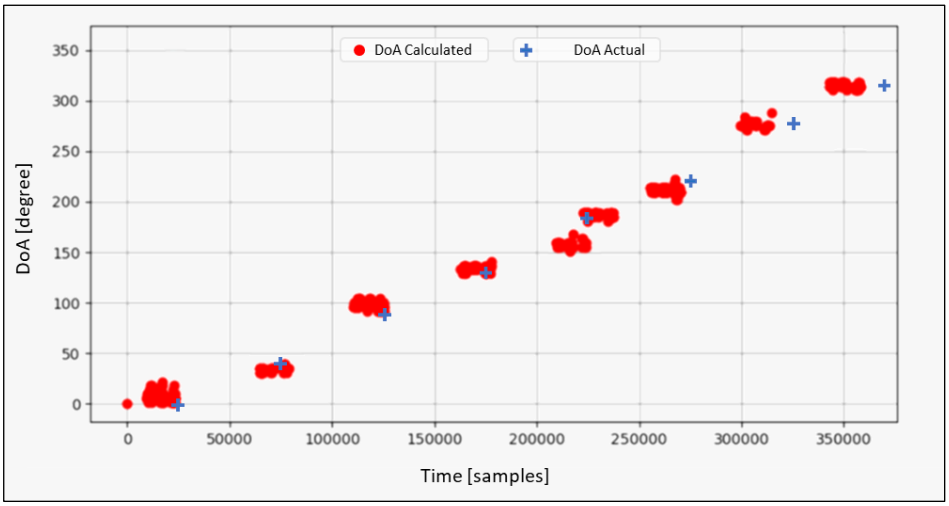

Figure 3.20. Calculated and actual direction of arrival.

Average displacement from correct position obtained from the figure is about ±4.5°. In order to improve resulting DoA, window size and correlation threshold parameters in the algorithm were tried to be fine-tuned. Effect of change in sampling rate was also evaluated. With the decrease of sampling rate from 44.1 kHz to 16 kHz, uncertainty of calculated DoA was increased about 1.8 times larger by keeping other parameters constant.

3.4.3. Beamforming Using Time Delay of Arrival Information

As a last step of this chapter, quality and level of localized sound data was improved by using simple beamforming method. Since all required parameters including delays between channels calculated, all signals were shifted accordingly to close delay gaps observed. Figure 3.18 shows original and corrected (delayed according to) waveforms of 4 channels separately.

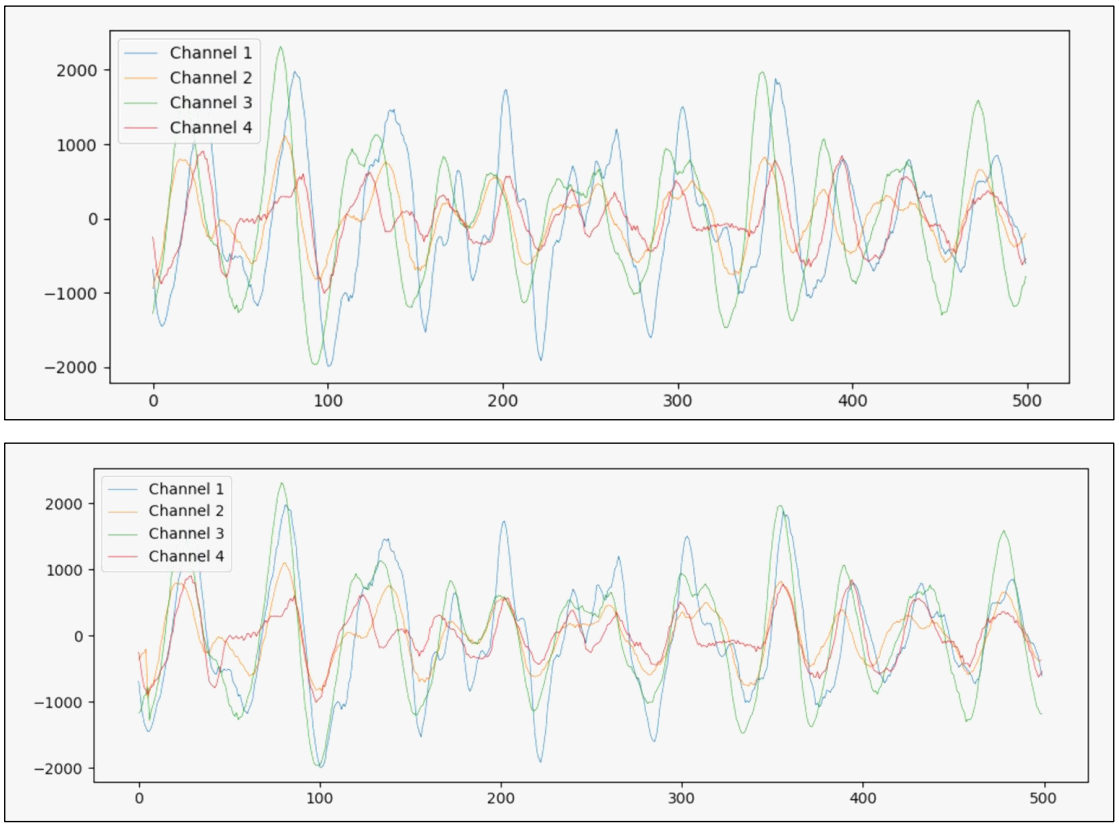

Figure 3.21 Original and Corrected (delayed) waveforms

Using delay and sum method mentioned in section 1.3.2, delayed channels were summed to obtain beamformed signal as shown in Figure 3.19, together with summation of original signal for comparison.

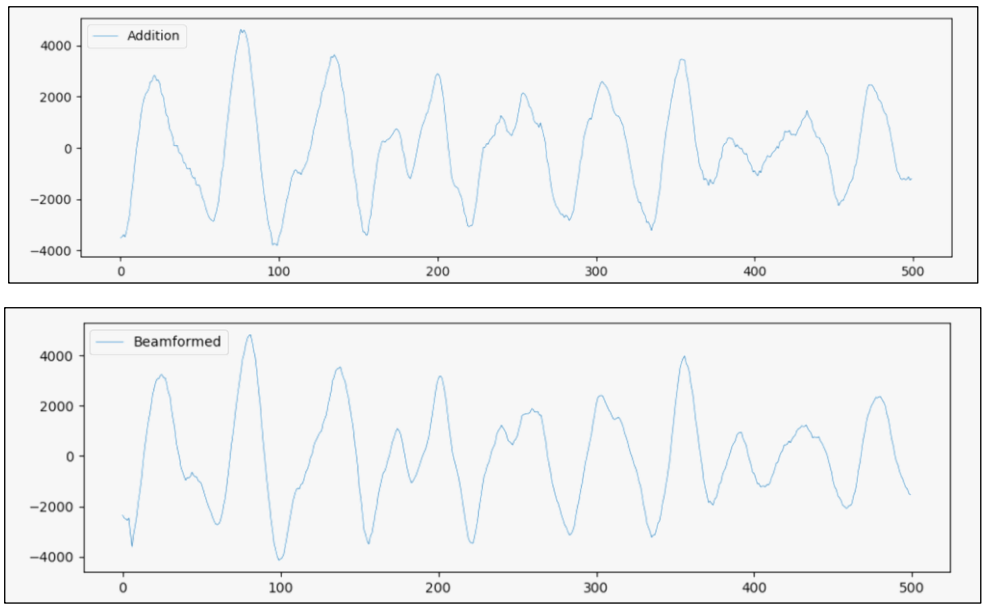

Figure 3.22 Addition of original signals and Beamforming result

After this step, noise components coming from other directions than DoA were mostly canceled-out due to delayed sums whereas desired component coming from DoA was amplified by 4 times. Beamformed signal is easily distinguishable with its smoother waveform and lower noise level than original one.

3. CONCLUSION

In this thesis study, end-to-end MEMS microphone array device development process were studied and practiced for sound source localization and beamforming applications. Through the study, theory of all corresponding components, and experiment results were discussed detailly.

As first step, working principles of MEMS microphones and MEMS microphone arrays were studied. During this phase of study, alternative geometries were evaluated together with existing commercial products to specify ideal array geometry to start with. To ensure alignment with sound source localization aim of study, time delay characteristics of sound signals were studied, and important constraints were defined.

Although there are different types of arrays for different applications, including many geometries in one array is now possible thanks to developments in digital MEMS microphone technology. Since no additional circuitry or peripheral is needed for communication of microphone and microprocessor, arrays containing 4+ microphones were defined as starting point of study.

Advantages, drawbacks and working principles of different digital sound interfaces used by MEMS microphones including PDM (Pulse Density Modulation), I2S (Inter-IC Sound) and TDM (Time-Division Multiplexing) were studied and compared. Thanks to the support for large number of microphones speaking simultaneously and other encouraging characteristics of TDM microphones, TDM interface was selected to be used in aimed array device. During this decision process, different microprocessor and evaluation boards alternatives were also reviewed to measure flexibility of each option before starting to implement.

During evaluation of TDM microphones, process of obtaining proper sound signal was failed a few times due to the various constraints and limitations. Since most of microcontrollers are not currently supporting TDM interface, different alternatives and workarounds were studied. Regarding high speed digital signal transmission and processing, different best practices such as spider network clock distribution were reviewed and used in prototype device.

Once successfully running microphone prototype array, one of the main objectives of thesis study was achieved. In next step, schematic design and PCB production steps were conducted carefully in parallel with experiments related sound source localization application.

Aimed device and experiment setup was designed to be compatible with a real-life environment. A keyword by male human was recorded and used in future parts of experiments as a reference signal and location of sound source tried to be calculated using different methods studied in previous parts. Delay between channels were dynamically calculated and this information was used both direction of arrival calculation and also beamforming operation.

As a result of all these processes, quality of the sound source which is desired to be recorded was increased using beamforming and location of this sound source was obtained with %97.5 accuracy (±4.5 degree over 360).

The works accomplished in this thesis have verified the convenience and reliability of usage of MEMS digital microphones in arrays. During practicing of relatively new interface, TDM, valuable insights and experience was obtained and shared through the all study. By concluding the study with production steps of own-PCB and device, all possible hidden limitations were observed, solved and device was successfully assembled.

The future work includes the combination of reliable MEMS TDM microphone arrays which was studied in the thesis, and advanced sound source localization algorithms supporting more than one data sources. By doing this, both cost-effective and high-performance sound source localization systems can be designed. Another improvement area to be studied is acoustic port design. Whereas omnidirectional characteristics of MEMS microphones are an advantage, shape of acoustic port and travel of sound waves are important areas to be studied and considered together with software algorithms in related scenarios.

REFERENCES

1. Sessler, G. M., and J. E. West. "Self- biased condenser microphone with high capacitance." *The Journal of the Acoustical Society of America* 34.11: 1787-1788., 1962.

2. Englewood, C., *Apple Products Are Driving Market Growth for MEMS Microphones, IHS Says.*, 2016, https://news.ihsmarkit.com/press-release/apple-products-are-driving-market-growth-mems-microphones-ihs-says, accessed in May 2019.

3. Sessler, G. M., "Silicon microphones.", *Journal of the Audio Engineering Society* 44.1/2 (1996): 16-22, 1996.

4. Elko, G. W., Pardo, F., López, D., Bishop, D., & Gammel, P., 2005, "Capacitive MEMS microphones" *Bell Labs Technical Journal*, 10(3), 187-198., 2005.

5. Walser, S., Loibl, M., Winter, M., Siegel, C., & Feiertag, G., *Sensitivity recalibration of MEMS microphones to compensate drift and environmental influences.*, Procedia Engineering, 168, 1759-1762., 2016.

6. Feiertag, Gregor, Matthias Winter, and Anton Leidl. "Packaging of MEMS microphones." *Smart Sensors, Actuators, and MEMS IV. Vol. 7362.*, International Society for Optics and Photonics, 2009.

7. Rombach, P., Müllenborn, M., Klein, U., & Rasmussen, K., "The first low voltage, low noise differential silicon microphone, technology development and measurement results.", Sensors and actuators A: Physical, 95(2-3), 196-201., 2002.

8. Martin, D. T., Kadirvel, K., Fox, R. M., Nishider, T., Liu, J., & Sheplak, M. , "Surface and bulk micromachined dual back-plate condenser microphone", 18th IEEE International Conference on Micro Electro Mechanical Systems, (pp. 319-322), 2005.

9. Dixon-Warren, S., *Survey of MEMS Microphone Technology*, 2016, http://www.chipworks.com/about-chipworks/overview/blog/survey-of-mems-microphone-technology, accessed in May 2019.

10. Widder, J., and A. Morcelli., *Basic principles of MEMS microphones*. , 2014, https://www.edn.com/design/analog/4430264/Basic-principles-of-MEMS-microphones-, accessed in April 2019.

11. Lewis, J. *Common Inter-IC Digital Interfaces for Audio Data Transfer*, 2012, https://www.analog.com/media/en/technical-documentation/technical-articles/MS-2275.pdf, accessed in April 2019.

12. Je, S. S., Kim, J., Kozicki, M. N., & Chae, "A directional capacitive MEMS microphone using nano-electrodeposits.", *IEEE 22nd International Conference on Micro Electro Mechanical Systems*, 2009.

13. Yoo, I, Sanghyeok Y., Hyunsoo K., "Development of the Directional MEMS Microphone for Automotive Applications", FISITA 2016 World Automotive Congress, Busan, Korea, 2016.

14. InvenSense Inc., *Microphone Array Beamforming*, 2013, https://www.invensense.com/wp-content/uploads/2015/02/Microphone-Array-Beamforming.pdf, accessed in May 2019.

15. Brown, P., *All the Technology Inside Your Google Home*, 2016, https://electronics360.globalspec.com/article/7718/all-the-technology-inside-your-google-home, accessed in May 2019.

16. IFixit.com, *Amazon Echo Teardown*, 2014, https://tr.ifixit.com/Teardown/Amazon+Echo+Teardown/33953, accessed in May 2019.

17. Seeed Studio, *ReSpeaker 4-Mic Array for Raspberry Pi*, 2017, http://wiki.seeedstudio.com/ReSpeaker_4_Mic_Array_for_Raspberry_Pi/, accessed in May 2019.

18. Valin, J.M., François M., and Jean R., "Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering.", *Robotics and Autonomous Systems 55.3 (2007): 216-228.*, 2007

19. Knapp, C. H. and Carter, G. C.: 1976, "The generalized correlation method for estimation of time delay, IEEE Transactions on Acoustics", *Speech and Signal Processing ASSP-24(4)*, 320-327, 1976.

20. Dmochowski, J., Jacob B., and Sofiène A., "On spatial aliasing in microphone arrays.", IEEE Transactions on Signal Processing 57.4 (2008): 1383-1395., 2018

21. Xiph.org, *24/192 Music Downloads are Very Silly Indeed*, 2013, https://people.xiph.org/~xiphmont/demo/neil-young.html, accessed in May 2019.

22. Kuri, V., Abir C., "Review on Sigma-Delta Modulator", International Journal of Engineering Research and General Science Volume 3, Issue 5, September-October 2015.

23. Kite, T., "Understanding PDM Digital Audio", *Audio precision (2012),* 2012

24. Philips Semiconductors , *I²S Specification*, 1986, https://www.sparkfun.com/datasheets/BreakoutBoards/I2SBUS.pdf, accessed in May 2019

25. Bosman, D., *VHDL i2s transmitter*, 2018, https://dwjbosman.github.io/vhdl-i-2-s-transmitter/, accessed in May 2019

26. Business Wire Inc., *InvenSense Announces World's First TDM Microphone With an Array of 16 Devices on a Single Bus*, 2016, https://www.businesswire.com/news/home/20160510006003/en/InvenSense-Announces-World%E2%80%99s-TDM-Microphone-Array-16, accessed in May 2019

27. InvenSense Inc., *ICS-52000: Low-Noise Microphone with TDM Digital Output*, 2017, https://www.mouser.com/pdfdocs/DS-000121-ICS-52000-v13.pdf, accessed in May 2019

28. Broadcom Corporation, *BCM2837 ARM Peripherals*, 2012, https://cs140e.sergio.bz/docs/BCM2837-ARM-Peripherals.pdf , accessed in May 2019

29. Atmel Inc., *SAM D21E / SAM D21G / SAM D21J ARM Based Microcontrolers Datasheet* , 2015, https://cdn.sparkfun.com/datasheets/Dev/Arduino/Boards/Atmel-42181-SAM-D21_Datasheet.pdf, accessed in June 2019

30. Arduino, *I2S Library*, 2016, https://www.arduino.cc/en/Reference/I2S, accessed in June 2019.

31. Atmel Inc., *8-bit AVR Microcontroller with 4/8/16/32K Bytes In-System Programmable Flash*, 2009, https://www.sparkfun.com/datasheets/Components/SMD/ATMega328.pdf, accessed in June 2019.

32. XMOS Inc, *xCORE Microphone Array Hardware Manual*, 2018, https://www.xmos.com/download/xCORE-Microphone-Array-hardware-manual(2V0).pdf, accessed in June 2019.

33. XMOS Inc., *VocalFusion XVF3000 mono-AEC voice processor*, 2017, https://www.xmos.com/products/voice/silicon#XVF3000-TQ128-C, accessed in June 2019.

34. Atmel Inc., *AT07451: SAM D21/DA1 Inter-IC Sound Controller (I2S) Driver*, 2015, http://ww1.microchip.com/downloads/en/AppNotes/Atmel-42255-SAM-Inter-IC-Sound-Controller-I2S-Driver_ApplicationNote_AT07451.pdf, accessed in June 2019.

35. Cirrus Logic Inc., *CS4244 - 4 In/4 Out Audio CODEC with PCM and TDM Interfaces*, 2014, https://statics.cirrus.com/pubs/proDatasheet/CS4244_F2.pdf, accessed in June 2019.

36. Invensense Inc., *ICS-43432: Low- Noise Microphone with I 2 S Digital Output*, 2015, https://www.invensense.com/wp-content/uploads/2015/02/ICS-43432-data-sheet-v1.3.pdf, accessed in June 2019.

37. Maxim Integrated Inc., *MAX9880A: Low-Power, High-Performance, Dual I²S, Stereo Audio Codec*, 2011, https://datasheets.maximintegrated.com/en/ds/MAX9880A.pdf, accessed in June 2019.

38. Manners, D., *Knowles, Infineon dominate MEMS microphone market*, 2014, https://www.electronicsweekly.com/news/business/finance/knowles-infineon-dominate-mems-microphone-market-2014-06/, accessed in June 2019.

39. Knowles Inc., *SPU0414HR5H-SB Amplified SiSonic Microphone*, 2012, http://www.mouser.com/ds/2/218/SPU0414HR5H-SB-28439.pdf, accessed in June 2019.

40. Knowles Inc., *SPH0645LM4H-B I2S Output Digital Microphone*, 2015, https://cdn-shop.adafruit.com/product-files/3421/i2S+Datasheet.PDF, accessed in June 2019.

41. NotWired.CO, *NW-AUD-ICS52000 TDK InvenSense ICS-52000 Low Noise Microphones*, 2017, https://www.notwired.co/datasheets/notwired/DS-NW-AUD-ICS52000.pdf, accessed in June 2019

42. Johnson, H. W., Martin G., *High-speed digital design: a handbook of black magic. Vol. 155.*, Upper Saddle River, NJ: Prentice Hall, Chapter 11, 1993.

43. Baken, R. J., and Robert F. O., *Clinical measurement of speech and voice*, Cengage Learning, 2000.

44. Brixen, E. B., *Facts About Speech Intelligibility*, 2016, https://www.dpamicrophones.com/mic-university/facts-about-speech-intelligibility, accessed in June 2019.

APPENDIX A: DEVICE AND EXPERIMENT PICTURES

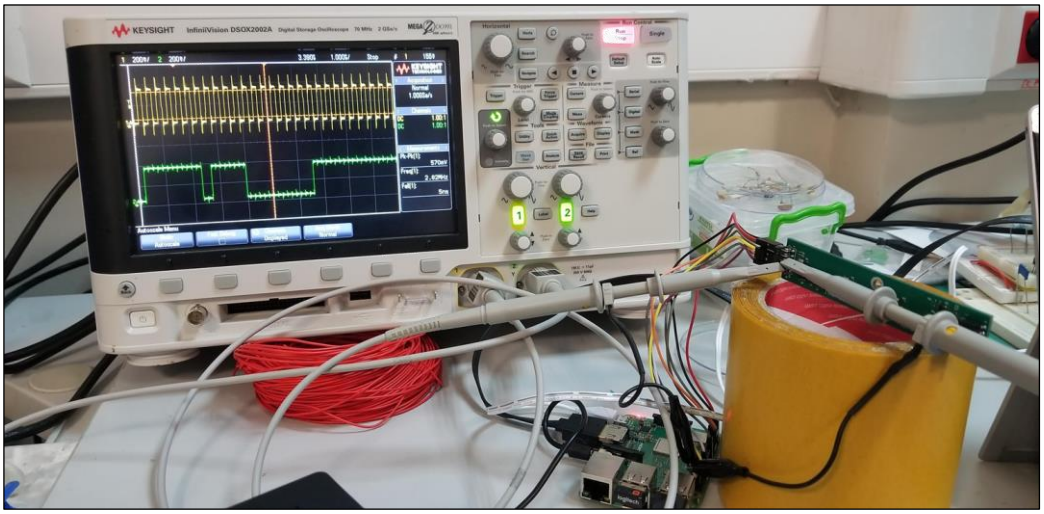

Figure A.1 Experiment setup including ICS-52000 array with Raspberry Pi 3b+.

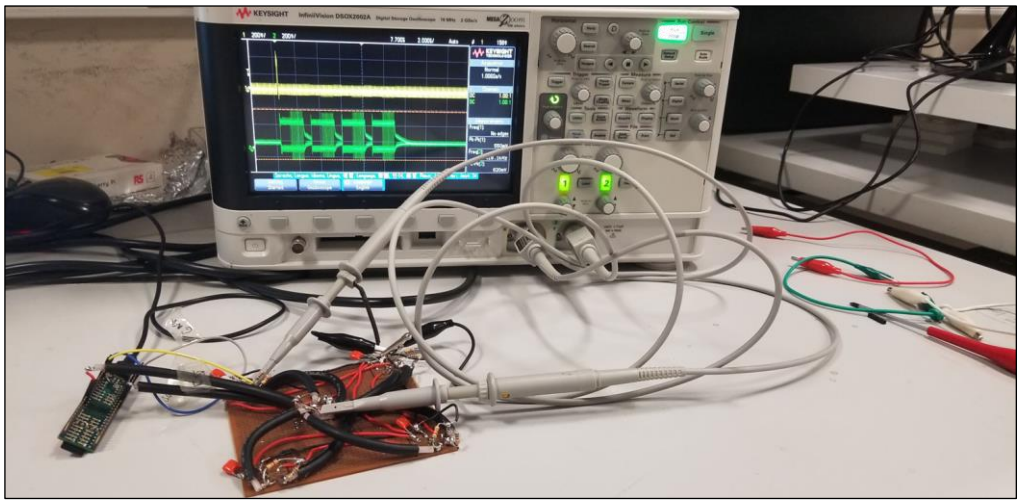

Figure A.2 Experiment setup including Teensy 3.2 and prototype device.

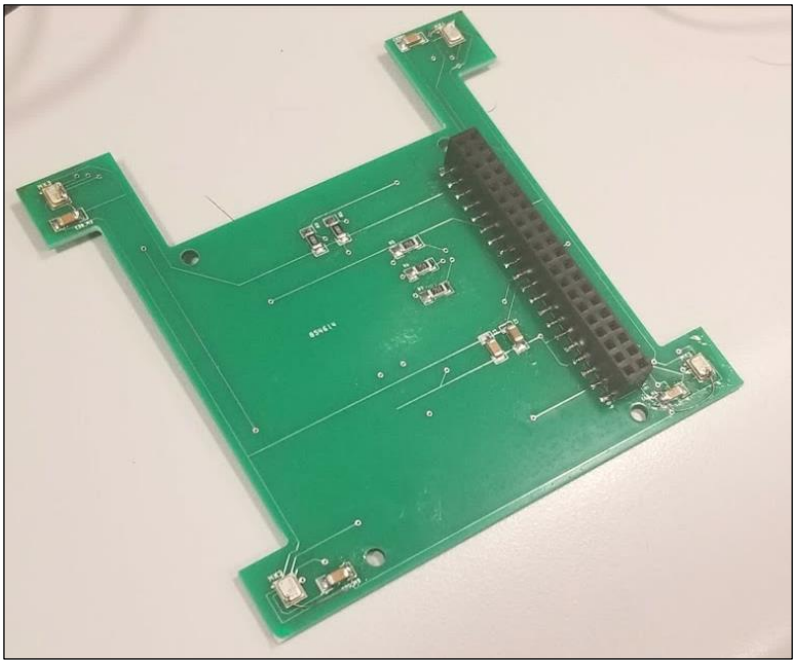

Figure A.3 Designed and produced PCB.

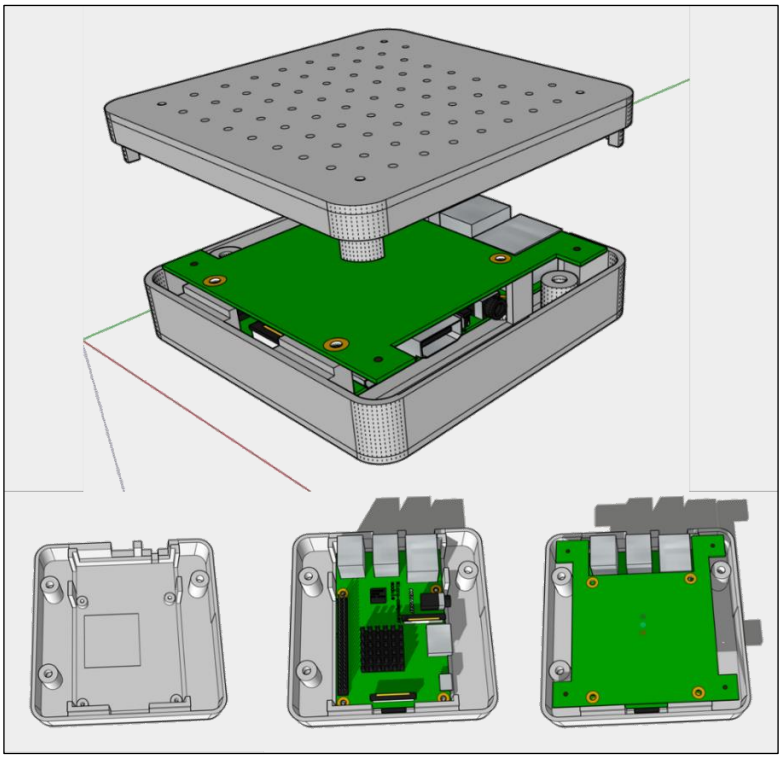

Figure A.4 Designed and printed device enclosure box.

Figure A.5 Assembled 3D printed device box.

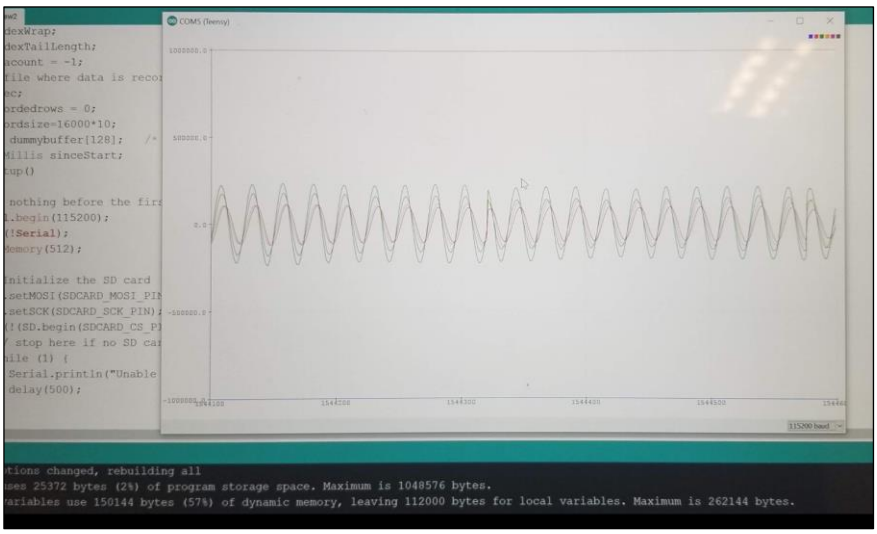

Figure A.6 Real-time recording of sinusoidal test waveform.

APPENDIX B: COPYRIGHT AND TRADEMARK INFORMATION

The implementation, analysis and verification of the digital microphone array required the variety of software and design tools.

The product or tool names and the company names are protected by copyright.

General copyrights and trademarks are acknowledged in this section to simplify further reading of the thesis.

Copyrights (© or ®) used in this thesis are:

- MATLAB
- Eagle Autodesk
- Sketch Up
- Audacity
- Solidworks

Trademarks (™) used in this thesis are:

- Python
- Audacity
- Raspberry Pi
- Arduino
- Teensy

Company names (inc and plc) used in this thesis are:

- Knowles Electronics/Acoustics
- InvenSense Inc.
- Seeed Studio
- Atmel Inc.
- NotWired Inc.
- XMOS Inc.