Final Report

for

Project 5 - Microphones for Stand Alone Voice Assistant

Version 1.0 Approved

Prepared by:

Marc Thibault - V00865176

Branden Voss - V00913539

Matt Hermary - V00848259

Jordan Pentney - V00907126

Group 9, Project 5

December 5, 2021

Table of Contents

1 Executive Summary

This report presents a conceptual analysis of group 9's proposed solution to the Microphones for Stand Alone Voice Assistant initiative under the project sponsor, Ergonomyx Technologies. The addition of such an accessory for Ergonomyx's new line of products, is being implemented to promote a physically healthier workplace experience [1]. The scope of work completed by group 9 entails microphone selection, microphone placement, microcontroller selection, implementation of a speech to text module, firmware applications and a system enclosure. The recommendations and feasibility of design choices made under the scope of this project are to be covered in this report. Findings of this analysis, in order to meet certain constraints outlined by the sponsor, suggest that the benefits of a single microphone system exceed that of implementing the beamforming concept. Furthermore, the system is to be implemented with the support of a microcontroller which resembles the capability of the ESP32 and has no network interface [2]. The motor control system for the desk is not covered in this report document.

This report is limited to a conceptual explanation of the microphone voice assistant system and its functionality. No physical prototype was realized and therefore, all recommendations and conclusions are based on focused research. The analysis of speech recognition capability and microphone functionality are also based on similar research projects and may not reflect the results obtained in these studies.

2 Introduction

In partial fulfillment for ECE 399, group 9 has been tasked with designing and assessing the feasibility of a stand-alone voice assistant under our sponsor, Ergonomyx Technologies. The voice assistant is intended to work in conjunction with Ergonomyx's new line of products which encourage users to lead healthier lifestyles in the workplace [1]. Current Ergonomyx workspace prototypes communicate to smartphones by using Bluetooth through a proprietary mobile app or home assistants such as Apple Siri, Google Assistant and Amazon Alexa [2]. Consequently, these assistants require data connection in a space where sensitive information may be conveyed. To overcome this, Ergonomyx has proposed developing an offline voice controlled system for the desk [2]. This document discusses a microphone system for a stand-alone voice assistant, a product intended to act as a voice controlled interface for Ergonomyx's smart sit and stand desk.

The main objective of this project is to gather information on microphones and propose the design for an offline stand alone voice assistant system for an office or home setting while maintaining as many of the existing features as possible [2]. The subsequent report is an integration of the literature review, requirements specification document and feasibility study conducted for the purpose of designing and analyzing the microphones for the stand alone voice assistant project.

3 Project Scope

Ergonomyx is developing a stand alone voice assistant to work in conjunction with their smart sit and stand desk. The voice assistant will allow users to control desk settings according to speech input received by a microphone configuration. The new system will replace the existing ESP32 microcontroller while working independently of data connection (i.e. no Wi-Fi or Bluetooth) and is to be fixed to the desk appropriately for the occupants use [2]. The system will be capable of receiving voice commands and making appropriate adjustments to the smart desk dependent on the given command. Several design evaluations have been made encompassing microphone selection, microphone configuration, additional hardware, firmware and enclosure specifications. For the proposed design, the main microprocessor to control the desk motors and other peripherals must be selected under project requirements to replace the existing ESP32 [3]. In conjunction with design specifications, an assessment of the practicality of the anticipated design was conducted.

Limitations to the scope of this project entail the microcontroller selection, enclosure specifications and minimum acceptable speech recognition percentage. The current microcontroller for the sit-stand-desk uses an ESP32, but these have Bluetooth and Wi-Fi capabilities which must not be used [3]. The new microcontroller will replace the ESP32, and needs to provide peripherals for 2 UARTs, 1 I2C, 1 ADC, 1 PWM and 5 digital IO pins [3]. In regards to the enclosure, IP50 and IP54 ratings must be met to protect against dust and splashing of liquids [3]. The enclosure is to be mounted under the desk, with most of the enclosure being within the desk [3]. Furthermore, a microphone configuration along with supporting hardware and firmware must be capable of offering at minimum, 60% speech recognition [3].

4 Literature Review

The literature review was developed after extensive research on each aspect of the microphone voice assistant module. This research was analysed as a group in order to obtain the most relevant information to use for further development of the project. The relevant aspects of the project was then broken down into the following sections:

- Analog vs Digital Microphone
- Standalone vs Array Microphone Arrangements
- Microphone Placement
- Speech Module IC’s
    - Acoustic Data Receiving Algorithms
- System Enclosure

Digital microphones were found to have better sound quality and less susceptible to interference caused by noise on the outgoing data line when compared to analog microphones [4]. The digital microphone achieves these better operating characteristics by using a MEMS transceiver rather than a standard diaphragm and voice coil [5]. However, analog microphones required less operating power and provided a smaller footprint compared to its digital counterpart [4].

When using either a digital or analog microphone in an array (2 or more microphone working in unison), special directional attenuation can be achieved through methods like beamforming [6]. Beamforming uses the inputs of multiple microphones (which all contain the same data but with slight phase differences) and puts them through custom designed signal adders to attenuate noise from unwanted directions. This allows for unwanted voices from certain directions to be neglected. When a standalone microphone is used, it presents a cheaper and easier design to incorporate into the final product, with little to no hardware interfacing issues with the speech module. The placement of the microphone(s) also plays an important role in reducing the amount of attenuation of the received signal. It was found that by placing the microphone as close as possible to any near flat surface, reflective interferences can be reduced, this in conjunction with proper microphone orientation (according to the microphone’s polar pickup pattern) can ensure effective voice recognition. The speech IC is then responsible for receiving the signal from the microphone(s), then sending the corresponding software defined output to the microcontroller, which will then execute the designated task.

There are numerous ICs for which speech to text can be easily obtained and almost all of them have some sort of GPIO and serial interfaces, which are to be used for interfacing with the microcontroller. The most prominent speech to text modules are the Elechouse V3 and Movi shield for Arduino. The Elechouse V3 has only 80 customizable commands but provides various GPIOs directly on the chip, whereas the Movi shield fits perfectly on an Arduino platform and allows for 150 custom commands [7, 8].

The speech to text modules already use the appropriate feature extraction algorithms for extracting voice commands such as Hidden Markov Models, Perceptual Linear Prediction, Linear Predictive Coding, and FFT [9, 10]. Therefore, there is no need to develop proprietary data receiving algorithms (as they are already implemented on the speech to text IC). Lastly, the entirety of the system must be encapsulated in some kind of enclosure. A small electronic enclosure must provide adequate chemical resistance and allow for necessary mounting and cable entries. Enclosures such as extruded aluminum, molded ABS, CNC machined, and even 3D printed enclosures were found to provide a potentially adequate solution, since the environment the enclosure will be exposed to is not very harsh. Metal enclosures typically offer more coatings and finishes than ABS and 3D printed enclosures. However, 3D printed enclosures offer fast and cost effective custom solutions for rapid prototyping.

5 Feasibility Study

The intent of the feasibility study was to conduct an analysis of the proposed project, the voice assistant microphone module for Ergonomyx’s sit stand desks, and the likelihood of this project meeting its goals as outlined by the sponsor company. Furthermore, it is meant to discuss and justify design decisions regarding the project's performance and cost.

5.1 Proposed Solution

The proposed system is shown below in figure 1. Details about each component and required configuration are covered in later sections.

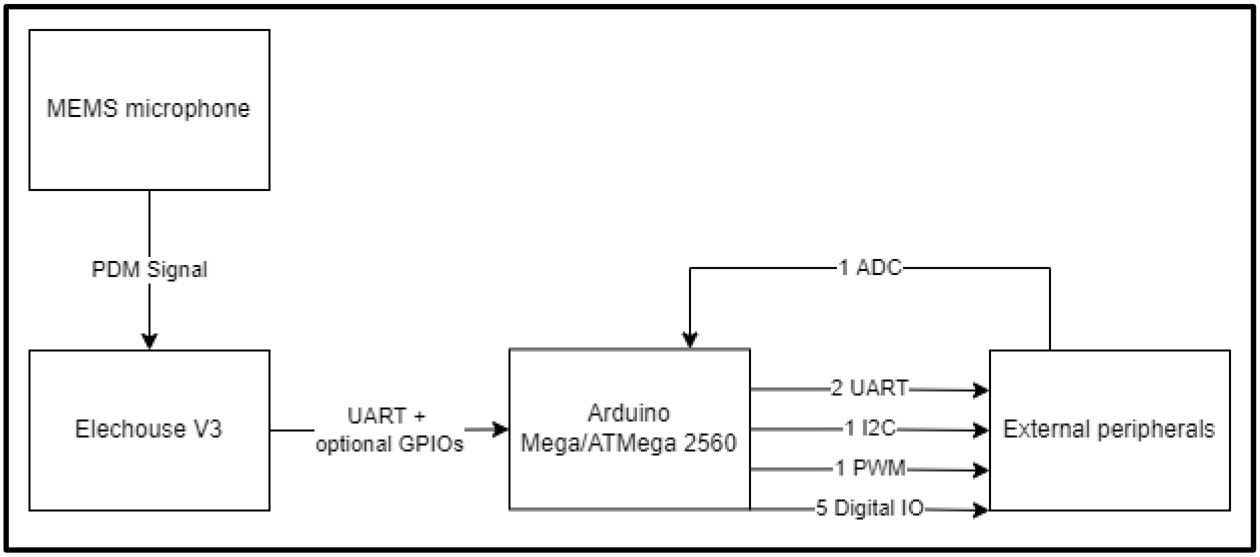

Figure 1: Overall design

5.2 Microphone Selection

A single Knowles SPC18P8LM4H-1-8 MEMS digital microphone is to be implemented in our design.

Electrical noise immunity

Sensitivity of -26dBFS and 120 maximum SPL

Cost effective at $1.01 [11]

5.3 Microphone Placement

The microphone is to be placed inside the enclosure underneath the desk near the side edge. This protects the enclosure from most hazards while not significantly blocking the microphone from incoming sound waves.

5.4 Speech to Text Module

Elechouse V3 speech to text module supports UART and digital I/O. This module is expected to exceed the requirement for 60% speech recognition, for according to its datasheet, it works with 99% accuracy in ideal environments, giving a significant buffer to work with. Additionally, it supports UART to pass specific messages to the main microcontroller upon certain phrases being heard. This was more versatile than the MOVI shield for Arduino UNO at a better price point.

5.5 Microcontroller

The Arduino Mega with the ATMega 2560 chip built in is our best option as it meets the requirement for no wireless data connection and resembles the ESP32 in its cost, capability and hardware interface features [12]. The ATMega has better hardware support and more similar features to the existing ESP32, which made it a suitable fit when compared to STM32 boards.

5.6 Firmware

Time required to port to a different microcontroller would depend largely on the complexity of the system. Fortunately the ESP32 can be programmed in Arduino. If this is the case for their existing firmware, the port will be fairly straightforward. Any existing code that relies on any form of internet connection to function (ex. MQTT, NTP, ect.) would need to be removed or have some alternatives designed for it.

5.7 Enclosure

The enclosure is to be made of aluminium to match the desk frame and will meet IP50/54 rating for dust protection and liquid spills as per project specifications [3].

5.8 Budget Estimation

Our budget review, taking into consideration the MEMS digital microphone, ATmega microcontroller, speech to text module and lastly the enclosure is as listed in the table below. At this time our budget is $23.66 excluding the price of the enclosure for which we will require a quote. For quantities of 10,000 the price came down slightly to $22.26 excluding the enclosure.

Component	Cost for 1	Per Unit Cost for 10,000
Knowles SPC18P8LM4H-1-8	$1.01* [11]	$1.01* [11]
Arduino Mega (ATmega 2560)	$5.95 [13]	$5.65 [13]
Elechouse V3	$16.70 [14]	$15.60 [14]
Enclosure	$x**	$x**
Total	$23.66	$22.26

*MEMS microphone must be ordered on a reel of 5,900

**Enclosure is highly variable based on required electrical components to be housed.

5.9 Feasibility

From the evaluation of the above budget breakdowns, this project is feasible for both prototyping and large production runs.

6 Evaluation

In order to evaluate our design, the components of our solution have been evaluated as follows.

6.1 Microphone

To determine if the microphone could accurately detect incoming voice commands, the audio sound pressure level (SPL) at the microphone had to be determined. While the user is seated or standing at their desk, we expect the distance between them and the microphone to be around 1 meter. For typical noise levels within an indoor setting, the speech level measured at a distance of 1 meter from the source comes in around 60 dB [15]. The sensitivity for digital microphones is measured in dBFS. The microphone we have selected has a sensitivity of -26 dBFS. The mathematical proof to verify if this sensitivity is sufficient for our solution can be seen in the calculations below[16].

$$Signal\ to\ Noise\ Ratio\ (SNR) = 94\ dB\ SPL\ reference\ signal\ -\ Noise\ Level$$
$$SNR\ =\ 34\ dB\ SPL$$
$$Equivalent\ Input\ Noise\ (EIN)\ =\ 94\ dB - \ SNR$$
$$Equivalent\ Input\ Noise\ (EIN)\ =\ 60\ dB$$

At a distance of 1 meter, -26 dBFS results in a maximum sound pressure level of 120 dB SPL. Therefore if the EIN were to go over 120 dB SPL the output of the microphone would be distorted. As calculated above our expected EIN is 60 dB. This puts the signal within the center of the sound pressure level range, giving a good buffer on either side to accommodate for noisier or quieter days in the office.

6.2 Voice Command Isolation Method

To ensure that voice commands from users are only recognized by that individual desk, a method to isolate users needed to be developed. Instead of implementing a beamforming solution which involves more microphones as well as further research and development, our solution uses custom activation words. By using custom activation words an individual's desk can be isolated before receiving a voice command. This allows us to meet the requirements while keeping our design simple and cost effective.

The minimum speech accuracy of 60% and existing hardware requirements are met by the Elechouse V3 and ATMega 2560 [3]. According to the Elechouse V3 data sheet, it is 99% accurate under an ideal environment [7]. However, the desk's environment is not ideal, therefore, a small deviation from 99% accuracy is expected.

6.3 Microphone Placement

The microphone placement was evaluated by considering a few placement options. Having the microphone directly underneath the desk would protect it the most against workplace hazards. However, in this position the microphone is also shielded from voice commands. To accommodate this, a hole could be drilled through the desk acting as a sound port. However, this results in the possibility of contaminants and hazards entering through the sound port. Due to this, we determined that the best placement for the microphone would be underneath the desk near the side edge. This placement keeps the microphone shielded from most hazards while decreasing how much the microphone is shielded from voice commands.

7 Conclusions and Recommendations

In conclusion, the microphones for standalone voice assistant project, sponsored by Ergonomyx has been studied and evaluated extensively which then lead to the conclusion that the design is in fact feasible. The recommended design uses a single MEMS analog output microphone (SPC18P8LM4H-1-8) in conjunction with an ATMega 2560 microcontroller. This design will allow a basic voice control system as well as have more than enough GPIO and serial interfaces to control the desk's existing functions. The ATMega 2560 chip is selected due to its large array of GPIO and serial interfaces, as well as having no on-board networking capabilities (security requirement from sponsor) [3]. The enclosure will be a basic bent aluminum 5-sided enclosure, finished to match the desk's frame components. The proposed system does, however, require some additional recommended work, mainly a quote for the folded aluminum enclosure, implementing the wake up word, designing a PCB for the speech module, and interfacing the MEMS microphone's 1-bit output data stream to the microcontroller. Overall, this solution should provide an enhanced solution compared to the current desk controller, and allow for the required voice command system.

8 References

[1] "Sit Stand Move," *Ergonomyx*. [Online]. Available: https://www.ergonomyx.com/. [Accessed: 19-Oct-2021].

[2] A. Branzan Albu and T. Porto Marques, *ECE 399 (11033) 2021 Projects*. University of Victoria, 2021, ch. Project 5: Microphones for Stand Alone Voice Assistant. [Accessed: 2-Dec-2021]

[3] Ergonomyx Technologies, "Questions for Client," Victoria BC, 2021[Accessed: 3-Dec-2021]

[4] S. Audio, "Why go digital: Analog vs digital wireless microphone?," *syncoaudio*, 30-Oct-2020. [Online]. Available: https://www.syncoaudio.com/blogs/news/why-go-digital-analog-vs-digital-wireless-microphone. [Accessed: 13-Oct-2021].

[5] "Analog and digital MEMS microphone design considerations." [Online]. Available: https://www.analog.com/media/en/technical-documentation/technical-articles/Analog-and-Digital-MEMS-Microphone-Design-Considerations-MS-2472.pdf. [Accessed: 13-Oct-2021].

[6] What is a microphone array? [Online]. Available: http://www.learningaboutelectronics.com/Articles/What-is-an-array-microphone. [Accessed: 13-Oct-2021].

[7] The Zen Cart™ Team and others, "Speak recognition, voice recognition module V3," Speak Recognition, Voice Recognition Module V3 [AD_VR3] - $28.00 : Elechouse, Arduino Play House. [Online]. Available: Speak Recognition, Voice Recognition Module V3 [AD_VR3] - $28.00 : Elechouse, Arduino Play House. Available: https://www.elechouse.com/elechouse/index.php?main_page=product_info&cPath=&products_id=2254. [Accessed: 16-Oct-2021].

[8] "Movi," Audeme. [Online]. Available: https://www.audeme.com/movi.html. [Accessed: 10-Oct-2021].

[9] P. Giannoulis, G. Potamianos, and P. Maragos, "Room-localized speech activity detection in multi-microphone smart homes," EURASIP Journal on Audio, Speech, and Music Processing, vol. 2019, no. 1, 2019. [Accessed: 03-Oct-2021].
[10] S. Adams, P. A. Beling, and R. Cogill, "Feature selection for Hidden Markov models and hidden semi-markov models," IEEE Access, vol. 4, pp. 1642–1657, 2016. [Accessed: 03-Oct-2021].

[11] "SPC18P8LM4H-1-8 knowles: Mouser," Mouser Electronics. [Online]. Available: https://www.mouser.ca/ProductDetail/Knowles/SPC18P8LM4H-1-8?qs=QNEnbhJQKvZpq1zSw13Qag%3D%3D. [Accessed: 30-Nov-2021].

[12] “Arduino board comparison guide,” Arrow.com, 28-Feb-2021. [Online]. Available: https://www.arrow.com/en/research-and-events/articles/arduino-uno-vs-mega-vs-micro. [Accessed: 20-Nov-2021].

[13] “Mega2560 R3 Atmega2560-16au Ch340g Avr Usb Board Development Board For Arduino,” Buy Mega2560,Atmega2560-16au,Usb Board Product on Alibaba.com. [Online]. Available: https://www.alibaba.com/product-detail/MEGA2560-R3-ATmega2560-16AU-CH340G-AVR_62388739627.html?spm=a2700.galleryofferlist.normal_offer.d_image.4adf72ceFk4ZWR. [Accessed: 30-Nov-2021].

[14] “Home,” *Easy-control Speak Recognition Voice Recognition Module V3 With Microphone - Buy Voice Recognition Development Board,Voice Module,Voice Command Module Product on Alibaba.com*. [Online]. Available: https://www.alibaba.com/product-detail/Easy-control-Speak-Recognition-Voice-Recognition_1700001536134.html?spm=a2700.galleryofferlist.normal_offer.d_image.127a60d0X1xmqA. [Accessed: 30-Nov-2021].

[15] “Required voice level at distance,” Engineering ToolBox. [Online]. Available: https://www.engineeringtoolbox.com/voice-level-d_938.html. [Accessed: 05-Dec-2021].

[16] “Microphone specifications explained - TDK.” [Online]. Available: https://invensense.tdk.com/wp-content/uploads/2015/02/AN-1112-v1.1.pdf. [Accessed: 05-Dec-2021].