Audio Engineering Society

Convention Paper

Presented at the 143rd Convention
2017 October 18–21, New York, NY, USA

A Low-Cost, High-Quality MEMS Ambisonic Microphone

Gabriel Zalles, Yigal Kamel, Ian Anderson, MingYang Lee, Chris Neil, Monique Henry, Spencer Cappiello, Charlie Mydlarz, Melody Baglione & Agnieszka Roginska

New York University, 35 W 4th St, New York, NY 10012

The Cooper Union for the Advancement of Science and Art, 41 Cooper Sq, New York, NY, 10003

Correspondence should be addressed to Gabriel Zalles (gz629@nyu.edu)

ABSTRACT

While public interest for technologies that produce and deliver immersive VR content has been growing, the price point for these tools has remained relatively high. This paper presents a low-cost, high-quality first-order ambisonics (FOA) microphone based on low-noise microelectromechanical systems (MEMS). Namely, this paper details the design, fabrication, and testing of a MEMS FOA microphone including its frequency and directivity response. To facilitate high resolution directivity response measurements, a low-cost, automatic rotating microphone mount using an Arduino was designed. The automatic control of this platform was integrated into an in-house acoustic measurement library built in MATLAB, allowing the user to generate polar plots at resolutions down to 1.8°. Subjective assessments compared the FOA mic prototype to commercially available FOA solutions at higher price points.

1 Introduction

With a currently expanding segment of the population becoming interested in Virtual Reality (VR), technology and media companies have begun investing in pairing visual experiences with truly immersive auditory content. While immersive audio used to be dominated by surround sound, interest in alternative immersive audio methods, such as Ambisonics, are expanding rapidly.

This paper presents a first-order ambisonics (FOA) microelectromechanical systems (MEMS)-based microphone. The project explores the effect of incorporating MEMS technology into a FOA recording and reproduction system. In particular, a FOA microphone was designed and built using MEMS capsules and a 3D-printed housing. A custom automatic rotating microphone mount (ARM^2) was designed and built to quantitatively test the directivity of the microphone at high resolutions. In addition, preliminary subjective testing was conducted to determine subjective preference between our prototype microphone and a commercially available solution from Sennheiser.

Many commercial FOA microphones, similar to the one proposed, have remained inaccessible to those outside exclusive University settings or professional-level studios. The goal is to develop a method for Universities and engineers to produce FOA mics independently and encourage people in general, through the process of building them, to become better acquainted with immersive audio technologies.

1.1 History of Ambisonics

Ambisonic technology was first explored in the 1970's by Michael Gerzon and Peter Fellgett [1]. Gerzon's work is based on the principle of spherical harmonics. By using four highly coincident capsules in a tetrahedral configuration, A-format signals can be encoded to a B-format matrix, which consists of three figure-eight pressure gradients and an omnidirectional pressure gradient, all coincidentally located. As noted by David Malham, the directional encoding of an ambisonic system is based on the ability of the spherical harmonics to approximate the surface of a sphere [2].

The FOA recording approach can also be considered an extension of the Mid-Side (M/S) technique created by the pioneer of stereophonic sound, Alan Blumlein, in the 30s [3]. The concept of decoding audio signals via a set of sums and differences is augmented in the FOA model which takes four signals and converts them into a zero-order information monophonic sound pressure component (W) and three first-order pressure gradients corresponding to the X, Y and Z axes, as shown in Figure 1.

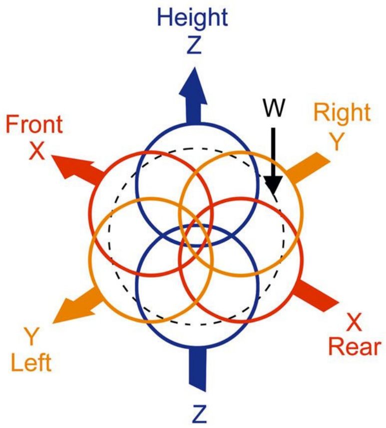

Figure 1: FOA encoded pressure gradients

Unlike other surround sound techniques, Ambisonics is isotropic, and as such, not speaker-centric [1]. Additionally, it does not rely on a specific speaker configuration. Rather, it decodes the localization data obtained in the recording process during playback to achieve a full 360-degree immersive representation of a soundfield.

1.2 MEMS Technology

In recent years, interest in MEMS microphones has expanded due to their versatile design, greater immunity to radio frequency interference (RFI) and electromagnetic interference (EMI), as well as low cost and environmental resiliency [4,5]. Current MEMS models are generally 10 times smaller than their more traditional electret counterparts. This miniaturization has allowed for additional circuitry, such as a preamp stage and an analog to digital converter (ADC), to output digitized audio, in some models, to be included within the MEMS enclosure. The production process used to manufacture these devices also provides an extremely high level of part-to-part consistency, making them more amenable to multi-capsule and multi-sensor arrays.

1.3 Capture and Reproduction

The most basic soundfield microphone, a FOA microphone, consists of four cardioid capsules mounted in a tetrahedral shape which captures a soundfield from a single point in space over four channels resulting in A-format signals, in their raw, unprocessed state. More complex systems which achieve higher-order ambisonics are possible by incorporating a greater number of capsules [2].

FOA relies on capsules with a cardioid response [1]. This aims to provide "acoustically transparent" sound capture; the signals captured from one microphone do not interfere with the signals from another capsule allowing for accurate and realistic reproduction of sound source location.

As mentioned, the A-format signals captured must be encoded to obtain one zero-order pressure gradient and three first-order spherical harmonics, each representing a different axis in three-dimensional space. The derivation of the B-format signals from the captured A-format is given:

$$\begin{aligned} W &= FLU + FRD + BLD + BRU \\ X &= FLU + FRD - BLD - BRU \\ Y &= FLU - FRD + BLD - BRU \\ Z &= FLU - FRD - BLD + BRU \end{aligned}$$

where FLU is Front Left Up, FRD is Front Right Down, BLD is Back Left Down and BRU is Back Right Up.

During decoding, these B-format vectors are projected onto either real or virtual speakers. In the case of virtual speakers, a head tracking system can also be incorporated to pan audio according to different listener head positions simulating the experience of sound radiating inwards from the surface of a sphere.

2 Design

The main considerations in the hardware design of this MEMS-based soundfield microphone were low cost, high audio quality, and the ability to capture faithful immersive sound fields. This section details key aspects of the design and construction of the prototype microphone including the implementation of a custom, automated rotating microphone mount for high resolution directivity measurements.

2.1 Capsule

The microphone technology chosen was the MEMS type, specifically the TDK InvenSense ICS-40720[1]. This specific capsule boasts a signal-to-noise ratio (SNR) of 70 dBA, acoustic overload point of 124 dB SPL, an unfiltered frequency response of 50Hz to 16kHz (±6 dB at response edges), and a low-noise differential output for reduced noise pickup over long cable runs. These microphones exhibit omni-directionality when operated without any coupled hardware such as a Printed Circuit Board (PCB) or housing.

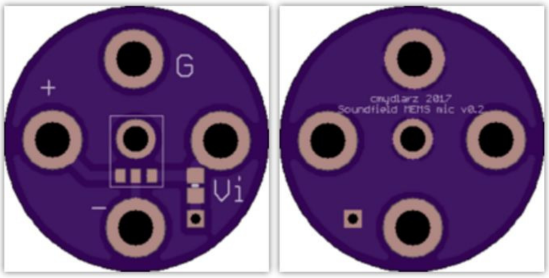

Figure 2: MEMS microphone board front side on right (12.5mm diameter)

The 12.5mm diameter PCB board is shown in Figure 2. A 0.1µF decoupling capacitor was placed between V_{IN} and GND to smooth out any power supply fluctuations that may result in parasitic noise in the audio signal chain.

Theoretically, it is possible to further reduce the size of these PCBs, further increasing the coincidence of the four transducers required for FOA recording, this will be further discussed in Section 5.

2.2 Housing Design

The housing for the MEMS-based FOA microphone prototype was 3D printed with Acrylonitrile Butadiene Styrene based filaments (ABS) using a high-end Stratasys Mojo 3D Printer[2], as shown in Figure 3. The housing serves not only to protect the capsules, but also to mechanically induce a cardioid response, as described in Section 3.2. The housing was designed according to the following specifications:

- The outer radius of the shaft of the housing was chosen to be 12.5mm so that the microphone could be mounted with a standard microphone clip.
- The inner radius of 8.5mm was chosen to allow sufficient space for the wiring while preserving the structural integrity of the shaft under pressure caused by mounting the microphone with a standard microphone clip.
- The head of the microphone was shaped to allow for each MEMS capsule to be positioned at the centroid of each face of a regular tetrahedron to allow for symmetrical soundfield acquisition.
- The reference tetrahedron used to position the capsules was chosen to have an edge length of 36mm, allowing sufficient space for wiring while keeping the overall design small enough to avoid significant spatial aliasing and associated distortion [1].

This spacing is consistent with similar FOA microphones on the market, where absolute coincidence is sacrificed in favor of practical considerations such as ease of fabrication and durability.

---

[1] invensense.com/products/analog/ics-40720

[2] stratasys.com/3d-printers/mojo

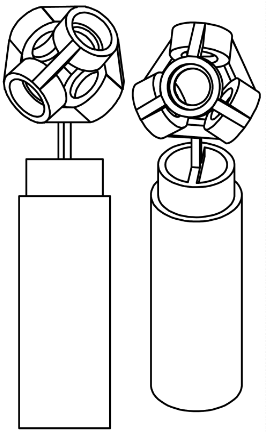

Figure 3: Custom microphone housing drawing

2.3 Automatic Rotating Microphone Mount (ARM²)

An automatic rotating microphone mount was designed in order to obtain the necessary polar response plots for the microphone. Manually measuring microphone directivity consumes considerable amount of time due to the inherent need to rotate the microphone some number of degrees repeatedly until at least 180° is reached for a single plot. Due to this necessity, automated rotating mounts are used to accurately and efficiently acquire the required data.

A rotating mount of this type was inaccessible due to high cost, therefore it was necessary to design a more cost-accessible system. Through the use of inexpensive parts, an automatic rotating mount capable of supporting both the MEMS microphone, and the Ambeo VR microphone, was designed, allowing comparisons between the two microphones to be made.

The ARM² was designed with the following requirements in mind:

- The mount was required to rotate along discrete steps with high accuracy and resolution in order to obtain accurate polar plots.
- The MEMS capsule was required to remain along the axis of rotation throughout the entire measurement process (see Figure 4).

To create the rotating mount according to the stated specifications, a NEMA23 3A stepper motor with “D-shaped” shaft was selected. The motor provides 200 steps per 360° rotation, resulting in a 1.8° rotational resolution. The motor’s high torque ensures the precision of each step under the load of the mount and microphone. The remaining specifications of the mount were accomplished through the use of a small microphone stand boom, a microphone clip, and a custom made stand attachment.

The attachment was used to secure the boom arm to the motor and was made by modifying an existing microphone stand. The boom arm allowed for telescoping and rotational freedom along a single axis while the microphone clip provided a second axis of rotational freedom. This mechanism enables the user to position the microphone at any angle while keeping the capsule centered directly above the motor’s axis of rotation. The complete ARM² construction is shown in Figure 4.

Finally, the in-house (NYU) developed Matlab application for directivity measurements, ScanIR [6], was extended to allow control of the stepper motor through an Arduino Uno Rev 3 and Adafruit Motor Shield V2. In conjunction, the ARM² and ScanIR generate polar plots of the microphone with high accuracy and efficiency at a low cost.

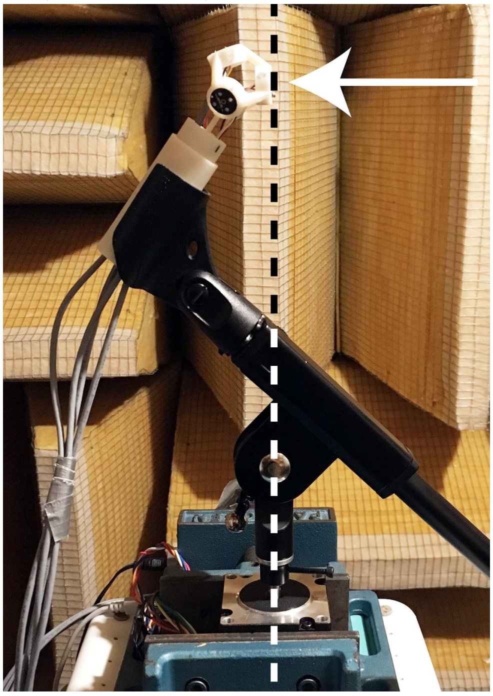

Figure 4: Automatic Rotating Microphone Mount (ARM²) with prototype microphone mounted in anechoic chamber at Cooper Union (microphone capsule marked with arrow at 0° incidence angle)

2.4 Cost Analysis

This paper focuses on a low-cost solution to FOA recording. During the design of the microphone, careful consideration was taken to minimize the cost of production.

Table 1: Approximate total microphone parts cost (excluding labor)

Component	Approx. cost (USD)
MEMS mics.	$8 (4 units)
MEMS PCBs	$5 (4 units)
3D Printed housing	$2
Interconnects	$10
Breakout board	$2
Total	**$27**

The parts used to manufacture this microphone and the ARM^2 rotating mount are all generally accessible to students and audio hobbyists. One of the main limitations of this solution would be the required skills to solder the microphone to its PCB and the accessibility of a 3D printer for the fabrication of the housing. Table 1 lists the approximate part costs for the prototype microphone excluding labor costs. These could be reduced with larger build runs.

Table 2: Approximate total cost for ARM^2 (excluding labor)

Component	Approx. cost
Arduino Uno	$25
Motor Shield	$20
Microphone Boom	$20
Stepper Motor	$20
12V Power Adapter	$2
Total	**$87**

The unit cost of $27 USD comes in well below the retail price of the Sennheiser Ambeo VR at $1650 USD as of July 2017, even when considering the markups involved on retail products including R&D labor and manufacturing costs. Furthermore, although not part of the microphone itself, the biggest financial burden avoided by design was that of the rotating platform for directivity tests. Equivalent systems[3] are inaccessible to those outside professional settings due to their high costs, sometimes in excess of thousands of USD. On the other hand, ARM^2 is far more accessible, with a total cost of under $100 USD.

3 Objective and Subjective Evaluation

This section details the testing procedure to profile the FOA microphone's acoustic characteristics, as well as subjective analysis methods. All objective measurements were carried out under anechoic conditions with ambient background levels of ~20 dBA.

3.1 Frequency Response

As previously discussed, the MATLAB toolbox, Scan IR [6], was used to generate the IRs of the microphones. The device under test (DUT), in this case the MEMS microphone, was compared to the professional grade Ambeo VR mic. Both of these microphones' frequency responses and polar plots were created using the swept sine method. The test signals were reproduced through a studio quality Mackie HR824 active speaker. A reference B&K 4189 microphone (assumed to be flat in frequency response from 20Hz-20kHz) was used to subtract the speaker's frequency response from the DUT's. Reference and DUT microphones were placed at 1m from the center point of the speaker on-axis, 1.3m from the anechoic chamber's floor.

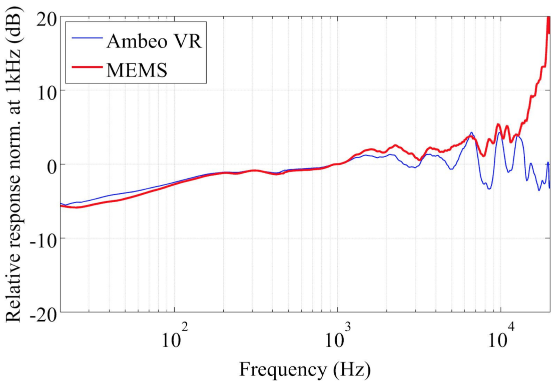

Figure 5: Single capsule frequency response of Ambeo VR & MEMS microphones

---

[3] nti-audio.com/Portals/0/data/en/NTi-Audio-Turntable-Product-Data.pdf

As can be seen in the Figure 5, both the MEMS microphone and the Ambeo VR microphone frequency responses are relatively flat between 100-12,000 Hz. Yet, whereas the peaks and troughs in sensitivity for the Ambeo VR mic, between 2-20 kHz, could be partially explained by possible passive filtering built into the audio signal's path, the MEMS microphone does not implement this circuitry. The rise in MEMS response after 10 kHz, is a result of the Helmholtz resonance created by the microphone's inner chamber [4]. The differences in subjective response to these characteristics will be discussed in Section 5.

3.2 Directionality

Using the ARM^2 rotating platform, the directivity of a single capsule from our MEMS prototype was measured. A single capsule of the Sennheiser Ambeo VR was also measured for comparison. Both microphones in question were measured with their respective capsules facing the speaker at a fixed distance of 1m. The ARM^2 was used to ensure that the capsule-speaker distance remains constant throughout the measurement cycle.

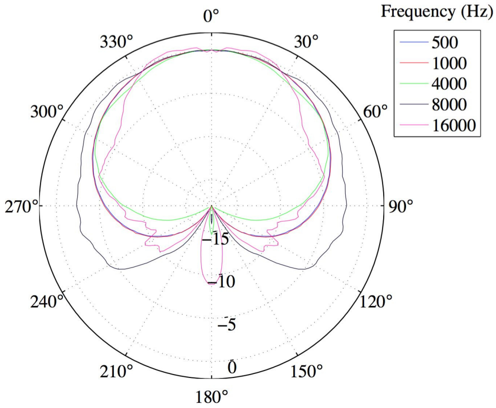

Figure 6: Mounted Ambeo VR capsule directivity at varying frequencies (rho values in dB)

As can be seen from Figure 6, the Sennheiser Ambeo VR microphone shows a clear cardioid polar pattern across all frequency ranges. This is likely due to the closed back construction of each capsule.

As shown in Figure 7, the MEMS capsule directivity exhibits a cardioid-like response at frequencies above 4kHz due to the effects of the housing and microphone PCB mount. The differences in perception between these two microphone recordings, when utilizing all four capsules and these have undergone B-format encoding, will be discussed in Section 4.

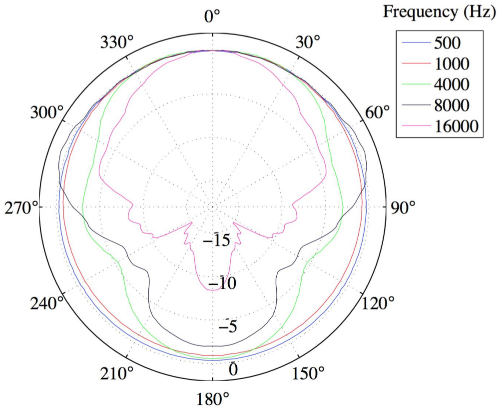

Figure 7: Mounted MEMS capsule directivity plot at varying frequencies (rho values in dB)

3.3 Subjective Evaluation

A preliminary subjective assessment was carried out using an online survey in order to determine the relative preference between the two recording solutions. Thirty-two participants were recruited from various university's music technology programs, audio-related mailing lists and small groups of non-audio experienced subjects.

The decoding of the B-format signals for reproduction was accomplished via the JavaScript Library, ForgeJS[4]. This library makes use of the binaural FOA decoder Omnitone written by Google using the Web Audio Application Programming Interface (API). This decoder passes subsets of the B-format signals (W, X, Y, Z) to eight virtual speakers arranged in a spherical configuration. By simulating the rotation and tilt of a listener's head, controlled via the subject's mouse or keyboard, or their phone's accelerometers and gyroscopes, subjects can rotate in virtual space. The decoder then provides scaling factors which dynamically modify the output of each speaker; this creates the sensation of being placed in the presence of the original soundfield. A set of Head Related Transfer Functions (HRTFs) are used in the final stage of the signal flow to binaurally encode the audio for accurate headphone reproduction.

Participants were instructed to use headphones as the ambisonic decoder would allow them to fully experience the immersive soundfield only if

---

[4] forgejs.org

experienced via headphones, since these do not corrupt the interaural-level and interaural-time difference information introduced by convolving the HRTFs with the virtual speaker's output during the decoding.

The thirty-two subjects were presented with two audiovisual experiences. The same visual content was used for both experienced. Each experience contained audio recorded with a different microphone; one was recorded using the MEMS microphone and the other using the Sennheiser Ambeo. Participants were not told which one was which. Subsequently, participants were asked to rate a number of metrics regarding the auditory experience in order to evaluate whether a noticeable degree of difference could be experienced between recordings.

The recordings were made in an acoustically treated audio research lab fitted with a set of Genelec 8030A speakers. The same audio was played back once for each of the two recordings, at the same level, over four loudspeakers; two of which were located in a stereo configuration, while a third was added in the center and a fourth was placed behind the microphones, which recorded the audio, one at a time. This configuration was chosen to present the audio at all sides of the microphones on a single plane to create a surround effect. Audio was recorded using a ZOOM F8 portable recording interface. Audio was then lined up, normalized, trimmed and faded in and out. The audio tracks were also normalized after the encoding process in MATLAB prior to exporting, in order to avoid clipping, particularly on the W channel.

Subjects were asked to give an estimate of their perceived level of experience with "music technology". This was done in order to determine a subjects' experience with microphone technologies and possible critical listening scenarios. An "expert" in music technology should be well versed in identifying different microphone types at varying price points and would therefore be more attentive in their evaluation of the differences between the two microphones. Subjects reported their levels of experience on a scale from one to five, with five being the highest score. The reported levels of experience are shown in Figure 8. Over 50% of participants self-reported a score of 4 or above.

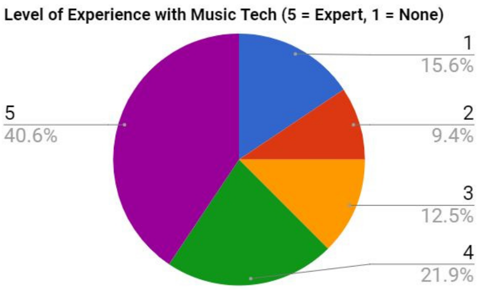

Figure 8. Breakdown of participants' reported level of music tech. experience during the subjective test

Subsequently, subjects were instructed to rate the two recordings, again on a scale from one to five, with five remaining the highest score, on the: naturalness, clearness, and accuracy of the reproduction, in that order. These questions were designed to uncover the microphones ability to capture: a realistic spatial environment (naturalness), a clear and detailed sound stage (clearness) and a timbrally accurate recording of the instruments (accuracy). Specifically, the questions were as follows:

"Does the performance appear to take place in an appropriate spatial environment?"

- 5 = Natural
- 1 = Unnatural

"Please rate from 1 to 5 how clearly the details of the performance can be perceived"

- 5 = Very clear
- 1 = Not clear at all

"Rate the accuracy of the timbral reproduction for both recordings"

- 5 = Very accurate
- 1 = Not accurate at all

Mean and standard deviation (Std) for the three questions are provided in Table 3. The results will be discussed in section 4.

Table 3: Mean scores and standard deviations (Std) for subjective ratings of the MEMS ambisonics microphone vs the Sennheiser Ambeo

	Ambeo		**MEMS**	
	Mean	*Std*	*Mean*	*Std*
Naturalness	3.38	1.16	2.94	1.16
Clearness	3.81	0.78	3.50	0.84
Accuracy	3.78	0.83	3.00	1.11

Participants were asked to select the recording they prefered overall, if any. Participants were also asked to optionally provide, in written form, a short comment on their opinions regarding the two recordings in open text response format. All data in the following analysis was submitted by subjects who were wearing headphones during the survey period.

4 Results and Discussion

Preliminary findings showed that subjects perceived a significant low-frequency reduction within the MEMS microphone recording, even though the measured low-frequency response showed little difference between the MEMS capsules and the Ambeo VR electret capsules. It is possible that the MEMS's omnidirectional polar response contributed to the subtraction of highly correlated low-frequency content during the encoding stage.

Subjects also noted that the MEMS recording contained overall more high frequency content than the Ambeo recording. While some noted a preference for this, others described it as overly bright. This observation is supported by the frequency response obtained for the MEMS capsules which show significantly greater dB levels than the Ambeo VR for frequencies above 2 kHz, especially above 10 kHz.

Furthermore, it was shown that the polar response of the MEMS microphone, whilst modified slightly by the microphone housing, had an overall negative impact on the subject's ability to perceive any panning compared to the Ambeo mic's recording. Subjects noted that this effect was particularly exacerbated when navigating the soundfield. The authors theorize that the superposition of signals, created by the low degree of directivity featured in each capsule, and its predominantly omnidirectional response across multiple frequency bands, was the main factor contributing to this phenomena.

Finally, it was found, from short comment responses, that just 12.5% of participants reported a higher noise floor in the MEMS recording, something which is often attributed to microphones containing very small capsules [4]. It should be noted, again, that over 50% of subjects in this study either considered themselves as semi-expert or expert with respect to music technology.

As shown in Figure 9, of the total valid number of subjects whose data was analyzed, 18.8% said they prefered the MEMS recording to the Ambeo VR mic. Of those, 50% reported a score of 3 or above in terms of their experience with music technology. 3.1% of the total population reported that the two recordings sounded the same. Overall, a preference towards the Ambeo mic can be seen as per the mean values shown in Table 3 for the three subjective criteria and the overall preference question. While these results were expected, results are promising considering the large price difference between our prototype solution and the Sennheiser Ambeo VR microphone.

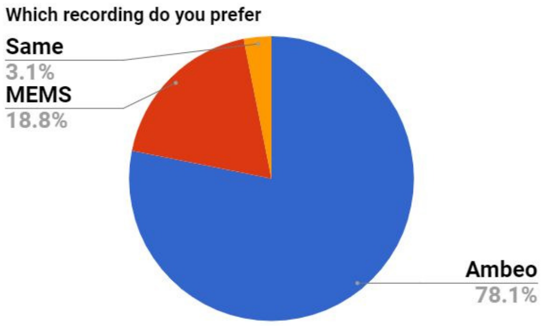

Figure 9. Breakdown of participants' responses to the question of which recording do they prefer

5 Future Work

A number of advancements to this prototype microphone could improve aspects of its physical design and overall functionality. Further experimentation on the effects of capsule directivity on subjective perception of FOA reproduction could result in the addition of backing materials applied behind the MEMS microphone capsules to mechanically induce a more cardioid response at varying frequencies. Uni-directional MEMS capsules could also be employed to enhance this effect. In addition, an investigation and evaluation into the effects of greater coincidence via smaller diameter PCBs may serve to enhance the preferred subjective response due to reduced spatial aliasing effects at high frequencies.

The structural integrity of the housing could be improved through a thicker connection between microphone shaft and head, or via fabrication using a different material. With the build and test procedure in place, the design, implementation, and testing of second and third order ambisonic MEMS based systems are also possible.

Finally, there is the necessity for a more elegant and dedicated powering system, likely through the use of a voltage regulator capable of bringing down Phantom Power (48V) to reasonable levels for MEMS operation. The prototype FOA MEMS microphone had been operating on a 3.5 V external battery, connected to a custom breakout board, located outside the housing of the mic. For this solution to appeal to consumers and audio purists, the circuitry within would have to conform to general studio conditions where phantom power is the preferred means of providing the microphone bias supply.

In addition, a comparative study with a comparable cost DIY ambisonic microphone solution such as the *SpHEAR Project[5] microphone may provide an additional benchmark for success.

6 Conclusions

The prototype MEMS-based ambisonics microphone shows promise in its ability to capture high quality 3D audio at a fraction of the cost of commercially available devices. Although the MEMS capsules operate remarkably well for their price and size, more research is necessary on the implementation of MEMS capsules and their use in immersive audio applications.

While the MEMS capsules directivity deviated from the desired cardioid response, its frequency and noise floor characteristics were generally well received. Results showed that subjects tended to perceive the MEMS recording as “thinner” and lacking bottom-end in general; however, most also noted that the MEMS capsules did not exhibit unfavorable signal-to-noise ratios.

The results showed the importance of using an ambisonic decoder, in our case Omnitone, during subjective assessment, for the audio quality evaluation. The ability for users to experience the movement of audio revealed that our mostly omnidirectional capsules failed to meet the criteria required for true FOA [1]. Without this 3D, online presentation medium, gauging the true differences in quality between the two mics would have been challenging.

The in-house designed ARM^2 system significantly reduced the time needed to carry out high resolution directivity measurements efficiently and accurately for the microphones under test. Despite subjective assessments indicating the strong preference for the Ambeo system, the MEMS FOA mic performed remarkably for a budget microphone using readily available hardware and software.

7 Acknowledgements

The authors would like to thank all those involved with this project, particularly the test subjects who helped gauge the subjective quality of the microphone, Sinisa Janjusevic from The Cooper Union for helping manufacture part of the ARM^2, Taylor Shield from NYU for helping prepare the CAD model for printing, and Paul Geluso for consulting on the mechanical induction of omnidirectional capsules and their effectiveness in an ambisonic context.

8 References

1. Gerzon MA. The Design of Precisely Coincident Microphone Arrays for Stereo and Surround Sound. Audio Engineering Society Convention 50. 1975.

2. Malham D. Higher Order Ambisonic Systems for the Spatialisation of Sound. International Computer Music Conference. International Computer Music Association; 1999. pp. 484–487.

3. Malham DG, Myatt A. 3-D Sound Spatialization using Ambisonic Techniques. Computer Music Journal. 1995. pp. 19-58.

4. Van Renterghem T, Thomas P, Dominguez F, Dauwe S, Touhafi A, Dhoedt B, et al. On the ability of consumer electronics microphones for environmental noise monitoring. J Environ Monit. 2011. pp. 544–552.

5. Mydlarz C, Salamon J, Bello JP. The implementation of low-cost urban acoustic monitoring devices. Applied Acoustics. 2017. pp. 207–218.

6. Boren B, Roginska A. Multichannel Impulse Response Measurement in Matlab. Audio Engineering Society Convention 131.

---

[5] cm-gitlab.stanford.edu/ambisonics/SpHEAR