Acoustic Imaging Using a 64-Node Microphone Array and Beamformer System

Feng Su
School of Information Technology
Carleton University, Ottawa, Canada
feng.su@carleton.ca

Chris Joslin
School of Information Technology
Carleton University, Ottawa, Canada
chris.joslin@carleton.ca

ABSTRACT

Acoustic imaging is difficult to achieve in noisy and reverberant environments. Microphone arrays offer an effective approach to obtaining a clean recording of desired acoustic signals in these environments. In this paper, we have designed, implemented, and evaluated a 64-node microphone array system for acoustic imaging. We have applied a delay-and-sum beamforming algorithm for sound source amplification in a noisy environment, and have explored the uses of the array and beamformer by generating the sound intensity map to reconstruct the acoustic scene of interest. Our experimental results show a mean error of 1.1 degrees for sound source localization, and a mean error of 13.1 degrees for source separation. In addition, we also used the system to image seven different materials with audible sound, and obtained their reconstructed acoustic maps as well as frequency response curves, from which we are able to detect the differences between textures based on their acoustic response powers.

Index Terms— Acoustic imaging, microphone array, beamforming, object detection, array signal processing

1. INTRODUCTION

Acoustic imaging can be described as a method for recording and reconstructing the amplitude distribution of a propagating sound field in a given plane. It is a field which has grown considerably over the past decade, and has been widely developed for important applications such as medical ultrasonography, non-destructive evaluation, and underwater sonar. While acoustic waves have been shown to be effective for imaging underwater and in the body, their use in-air is difficult since the propagation speed of sound in air is relatively low and interference from noise and echo is high. So far, a large amount of work has been done to explore possible solutions for in-air acoustic imaging.

More recently, new techniques have been developed to record and generate acoustic images in air. Microphone arrays are capable of providing spatial information for incoming acoustic waves, as they can capture key information that would be impossible to acquire with single microphones. Acoustic imaging microphone arrays often contain a camera which is usually located at the center of the array. An acoustic map, generated using the microphone data, is overlaid as a transparency over the camera image. With certain array signal processing techniques it is possible to localize individual sound sources in the recorded sound field and their emitted sound pressure level (SPL) from that acoustic image.

While a wide frequency range is available for acoustic imaging, ultrasound becomes one of the most widely used imaging technologies as it offers a much higher frequency (usually above 20 kHz) and can easily penetrate opaque media. This is why it has been successfully applied for imaging in human body and underwater sonar applications. However, ultrasound can be disrupted by air or gas, and therefore is not the most ideal imaging technique for applications that require to be operated in air. This has led us to explore other imaging modalities, such as using audible sound.

1.1. Contributions

Acoustic imaging with microphone arrays has been an ongoing topic for more than a decade. Compared to other array systems presented in the current literature, we highlight the following key characteristics of our system that contribute to its capability in acoustic imaging applications:

- High accuracy: Our system is capable of localizing sound source with an average error of 1.1 degrees. It can also separate two sound sources with an average error of 13.1 degrees, which achieves significant improvement over previous small-scale microphone array localization system.
- Low power consumption: All of the microphones are powered by a 3.3V/0.014A DC power supply, which is 46.2 mW in total. The output power of the speaker is 200 mW. Thus, the overall system power consumption is 246.2 mW. Compared to optic-based system such as Kinect (whose power demand is around 12 W), our system consumes approximately 98% less power.

- Robustness in the presence of noise and reverberation: Our array system can accurately locate and separate sound source signals in a noisy and reverberant environment, which proves its suitability for practical applications in everyday surroundings.
- Cost effectiveness: The overall cost for the system is around $630, or $1.745/cm^2 installed, which is relatively cheaper, especially compared to commercial products.
- Simple, efficient, easy to build and use: Our system is constructed entirely using commodity, off-the-shelf audio modules and 3D printing technology. The programs we developed allow for immediate visualization of the raw data as well as the obtained acoustic images.

We also examine the potential of our imaging array system for detecting objects with different textures. We obtained the acoustic images of 6 different materials and a human hand, and compared their frequency responses from 1 kHz to 7 kHz. To our knowledge, few works have ever measured such frequency responses within the range of audible sound, let alone in the context of acoustic imaging. From the results, we observed that while materials with smooth textures have similar frequency response patterns, the responses from rough textures (such as cardboard and human skin) present some significant differences at certain frequencies. Compared to smooth textures, their response powers are relatively weak from 3.5 kHz to 6.5 kHz. This demonstrates our system's effectiveness of detecting the differences between textures. We show that audible sound can be a cheap, low-power alternative technology for acoustic imaging.

2. RELATED WORKS

Over the past two decades, microphone array technologies have been increasingly applied for sound source amplification in difficult acoustic environments. Microphone arrays can be steered in software toward a desired sound source, filtering out undesired sources. When an appropriate level of computational power is available, microphone arrays can also track a desired source around a space as the source moves.

2.1. Sound Source Localization and Separation

One of the major functionality of microphone array signal processing is the estimation of the location from which a source signal originates. It is accomplished by utilizing differences in the sound signals received at different observation points to estimate the direction and eventually the actual location of the sound source. Extensive research on localization strategies exists. Pei et al. [1] presented an approach for locating a sound source using the small-scale linear microphone array on Kinect. Their positioning results showed an average error of 0.25 meters along the horizontal axis and 0.53 meters error along the vertical axis. Goseki et al. [2, 3] proposed a method of visualizing sound pressure distribution by combining microphone array processing with camera image processing. O'Donovan et al. [4] showed through a spherical microphone array that the passive localization of sound sources and their reflections in a concert hall is possible. Legg and Bradley [5] presented a calibration technique for an acoustic imaging spherical array with 72 microphones, combined with a digital camera. Their proposed technique obtained a mean position difference of 6.6 mm of the sound source coordinates in the acoustic maps.

In source separation with multiple microphones, the problem here is to separate different signals coming simultaneously from different directions. All the approaches are blind in nature since there is not usually access to either the acoustic channels or the source signals. Independent component analysis (ICA) [6] is the most widely used tool for the blind source separation (BSS) problem, since it takes fully advantage of the independence of the source signals. For example, Turqueti et al. [7] provided the first results on the use of a 52 microphone micro-electro-mechanical systems (MEMS) array, embedded in a field-programmable gate array (FPGA) platform as a source separation system utilizing the ICA technique. Similarly, Kajbaf and Ghassemian [8] proposed a new imaging method for heart sound segmentation, which was based on a smaller 3 by 3 microphone array using the delay-and-sum beamforming technique.

2.2. Object Detection

Object detection is commonly based on optical imaging sensors. For example, LIDAR and Kinect use infrared light, while stereo cameras use visible light. These systems require hardware operating at high sampling frequencies, precise calibration, and they dissipate significant power. Object detection by sonic or ultrasonic means is attractive because of its relatively low power consumption, and simpler, low-rate hardware. Additionally, it could complement light-based detection in scenarios where light fails, such as mirrors, windows, or spaces filled with smoke.

Moebus and Zoubir [9] studied ultrasound imaging in air for object detection, and discussed its suitability for biometric applications such as human presence detection [10]. Their system was based on beamforming with a synthetic 2D array of 400 acoustic receivers. Dokmanic and Tashev [11] designed a simple ultrasonic device with eight MEMS microphones and eight piezo transducers operating at 40 kHz, for acquiring images in both azimuth and elevation. They obtained depth images that revealed the pose of a human subject. This suggested that ultrasound could be used for skeletal tracking. However, due to its high attenuation nature, the use of ultrasound in air, especially in the presence of noise and reverberation, is still limited.

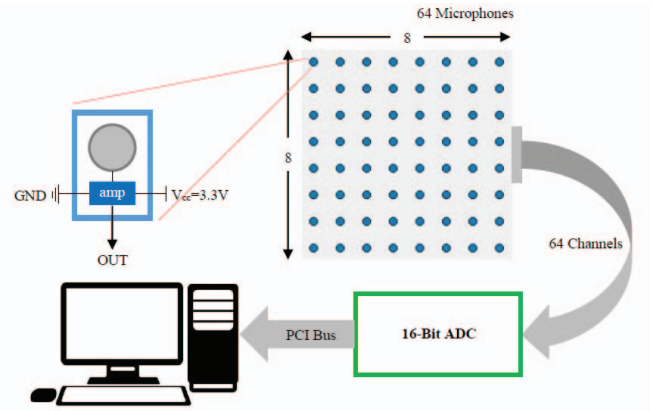

Fig. 1. The schematic diagram of the 64-node microphone array hardware design.

2.3. Array Signal Processing

Microphone arrays present additional challenges for signal processing methods since large numbers of detecting elements and sensors generate large amounts of data to be processed. Furthermore, applications such as sound source localization and sound imaging, require complex algorithms to properly process the raw data [12]. Many signal processing techniques that use arrays of sensors have been proposed to improve the quality of the output signal and achieve a substantial improvement in the signal-to-noise ratio (SNR). The most well-known general class of array processing methods are beamforming [13], which has already been widely used for many decades in different application fields, such as the Sound Navigation and Ranging (SONAR), Radio Detection and Ranging (RADAR), and ultrasound imaging. The beamforming technique requires the utilization of microphone arrays that capture all emanating sounds. All incoming signals are then combined to amplify the primary source signal, while at the same time suppressing any environmental noise.

3. DELAY-AND-SUM BEAMFORMING

Beamforming technology alleviates the majority of the shortcomings that other recording techniques introduce, at the cost of an increased number of input channels. A beamformer is a processor used in conjunction with an array of sensors to provide a versatile form of spatial filtering [13]. The sensor array collects spatial samples of propagating wave fields, which are processed by the beamformer. The objective is to estimate the signal arriving from a desired direction in the presence of noise and interfering signals. The beamformer performs spatial filtering to separate signals that have overlapping frequency content but originate from different spatial locations. The spatial-filter based beamformer was developed for narrowband signals that can be sufficiently characterized by a single frequency. It can be used for plenty of different purposes, such as detecting the presence of a signal, estimating the direction of arrival (DOA), and enhancing a desired signal from its measurements corrupted by noise, competing sources, and reverberation.

Currently, we are using a delay-and-sum beamformer [14], which is the simplest way of computing the beam. Delay-and-sum beamforming (DSBF) uses the fact that the delay for the sound wave to propagate from one microphone in the array to the next can be empirically measured or calculated from the array geometry. This delay is different for each direction of sound propagation, i.e., from the sound source position. By delaying the signal from each microphone by an amount of time corresponding to the direction of propagation and then summing the delayed signals, we selectively amplify the sound coming from a particular direction (called the look direction). DSBF assumes that the position of the desired source relative to the array is known.

In array design, we would expect to set the spacing among sensors as large as possible, which would, in general, lead to more noise reduction. However, when the spacing is larger than $\lambda/2 = c/(2f)$, where λ is the wavelength of the signal, c is the speed of sound, and f is the frequency of the signal, spatial aliasing would arise [12]. This would cause ambiguity in recovering the desired signal. On the other hand, the sensors cannot be too close. If they are too close, the array does not provide enough aperture for recovering the source signal. A general rule of thumb is to choose the spacing of sensors between $\lambda/10$ and $\lambda/2$.

4. SYSTEM IMPLEMENTATION

The design of the microphone array faces several challenges such as number of detectors, array geometry, interference issues and signal processing. These factors are crucial to the construction of a reliable and effective microphone array system. Fig. 1 illustrates the hardware design of the system. Our array is constructed using 64 omni-directional microphone modules whose operating frequency range from 20 Hz to 20 kHz. Each module consists of an electret condenser microphone (CUI CMA-4544PF-W) and an operational amplifier (Maxim MAX4466) with adjustable gain. These electret microphones are fundamental parts of the array due to their small size, high sensitivity, and low power consumption. When combined into an array, they provide a highly versatile acoustic aperture.

While many geometrical configurations of the array are possible and potentially desirable, our microphone array consists of 64 elements, and is laid out in a rectangular grid pattern, with 8 columns and 8 rows. This allows us to steer the amplification beam horizontally as well as vertically. In order to avoid spatial aliasing and have a good acoustic aperture, the inter-microphone distance is determined to be 2.3 cm center to center, and is maintained in both the horizontal and vertical directions. This decision was due both to practicality reasons, as well as preliminary experiments with various spacings. The 2.3 cm spacing is

approximated by dividing the sound speed by the highest signal frequency and then dividing the result by two in order to satisfy the Nyquist-Shannon sampling theorem. This spacing makes it possible to obtain relevant phase information of incoming acoustic sound waves, increasing the array sensitivity and allowing spatial sampling of frequencies up to 7400 Hz without aliasing.

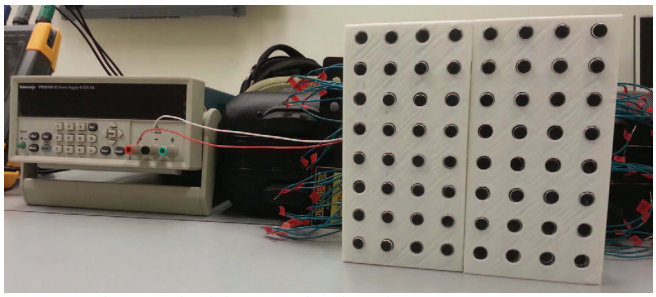

Fig. 2. A photograph of the 64-node microphone array system.

The body of the microphone array was 3D printed using currently available 3D printing technology. This allowed all the 64 microphones to be accurately positioned, maintaining a spacing of 2.3 cm in both the vertical and horizontal directions (Fig. 2). For the best performance, all the microphones were powered from a 3.3V and 0.014A DC power supply, and all the output pins were connected to the PMC66-16AI64SSA 16-bit data acquisition (DAQ) board through a Mini D Ribbon (MDR) cable. All the 64 channels were sampled simultaneously, and their outputs would pass through a digital processor which applied gain and offset correction values obtained during auto-calibration. After the audio data were transferred to the PCI bus, a computer connected to the DAQ board was capable of retrieving the data using the 16AI64SSA driver application program interface (API) software. This software also contained an auto-calibration function that calibrated all analog input channels to a single internal voltage reference in order to obtain maximum measurement accuracy.

The system presented in this work was set to work at 50k samples/second which was well above the Nyquist rate to avoid aliasing. The data collection time was 1 second for all of the 64 channels. Once the sampling rate and time were set, the system went into data acquisition mode, where the boards were continuously sending data to the computer and the computer piped this data to a binary file on the hard disk. After the collection, we created a data logger to import this file so that we could visualize the raw data on the screen, and further process and analyze the data with the algorithms we developed for the system.

According to the array geometry, signal operating frequency and propagation speed, we constructed a delay-and-sum beamformer and aimed it by appropriately delaying, attenuating, and adding the signals from each microphone so that they were in phase for one or more specific spatial locations. To utilize the beamformer to steer and scan the entire region in front of the array, we updated the look direction every time after the beamforming operation was performed so that it could scan the region from -90 degree to 90 degree in both azimuth and elevation. In addition, we also performed a fast Fourier transform (FFT) on the beamformed signal so as to bandpass filter the result to further reduce the low-frequency noise, and extract the signal containing the frequency of the sound source to acquire its relative intensity level. Finally, with these signal intensities at every direction in front of the array, we were able to generate the acoustic image and reconstruct the sound field.

Before conducting any experiment, it is necessary to calibrate the system. By adjusting the small trimmer pot on the back of each microphone module, we were able to calibrate the individual gain of each channel on the microphone array so that it could respond homogeneously when excited.

5. RESULTS AND DISCUSSION

5.1. Sound Source Localization

The localization experiments involved recording a moving sound source in a room where reverberation and several sources of noise were present. A wooden board (48 cm × 41.5 cm) with an 8×8 grid in the middle was placed 30 cm in front of the array. Each cell in the grid was 2.3 cm by 2.3 cm in order to match the spacing of the array. A speaker, 30 mm in diameter, producing continuous sinusoidal waves of 5 kHz was placed in one of the cells on the wooden board. This ensured that the speaker could be accurately positioned at every direction in front of the rectangular array. The speaker was moved between every cell in the grid while the tone was recorded by the microphone array. Each recording was 1 second long and was sampled at 50 kHz.

Fig. 3 (a) displays the results of the sound source localization at 64 different positions. We only tested 5 kHz for this experiment. The source is here localized using the sound intensity distribution map generated by the method described in Section 4. From each of the intensity map, we can clearly identify the location information of the sound source. To calculate the position errors, we measured the coordinate of the peak value from each of the sound intensity map, and used this information as the estimated source location. By calculating the distances between the coordinates at the four corners, and mapping these distances onto an 8 by 8 grid with a spacing of 5 degrees (or 2.3 cm), we were able to convert the coordinates acquired from Fig. 3 (a) to the same local coordinate system used by the real positions of the sound source. Compared to their real locations, the mean error of the estimated positions is 1.1 degrees (or 0.49 cm) with a maximum error of 3.1 degrees (or 1.44 cm). This demonstrates that our microphone array system has the capability of accurately estimating the location of the sound source.

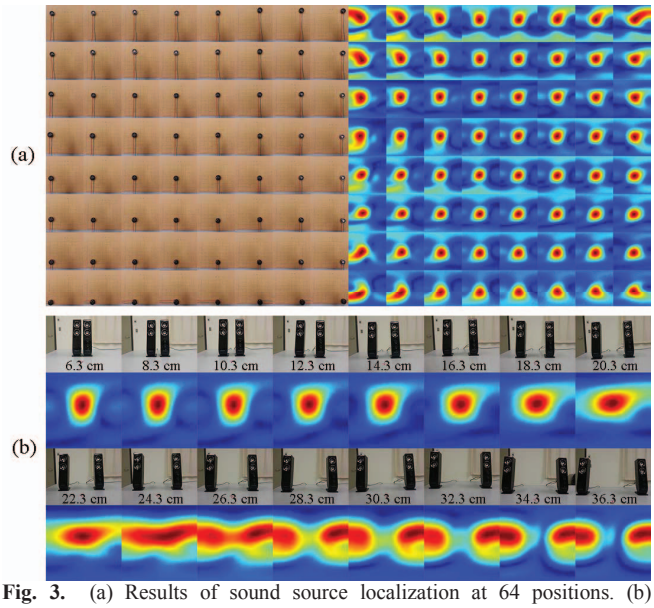

Fig. 3. (a) Results of sound source localization at 64 positions. (b) Separation of two sound source signals.

5.2. Sound Source Separation

Another case study to demonstrate the array system's performance is illustrated by Fig. 3 (b). In this experiment, the sound sources were two omnidirectional speakers (3.2 cm in diameter) emitting a single tone of 5 kHz at the same time and located 40 cm away from the microphone array. The spacing between the two sound sources was set to be 6.3 cm in the beginning, and then increased by 2 cm after each recording. Similar to source localization, all 64 channels of audio signals were sampled at 50 kHz for 1 second, and then processed by the DS beamformer in order to measure the sound intensities at every direction in front of the array. The results are shown in Fig. 3 (b). We see that the two source signals are mixed in the resulting image when the spacing between the speakers is less than 24 cm, and after this distance, the separated source images are becoming identifiable, allowing us to locate their relative positions. Using similar approach described in source localization, a mean separation error of 13.1 degrees (or 1.31 cm) was obtained for this experiment. This was based on the comparison of the estimated and real coordinates after the two separated signals were observed.

5.3. Imaging of Different Materials

To generate the acoustic images and test the frequency responses of different materials, we illuminated the object to be analyzed by a single sound source standing at a fixed position near the array. By this, we were able to image the scene by processing the back-scattered reflections from the object and analyze its acoustic response for a range of frequencies.

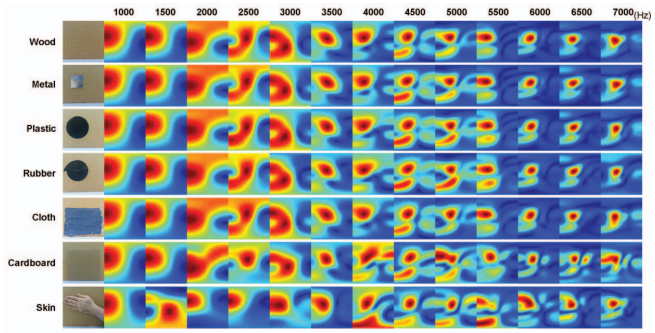

Fig. 4. Reconstructed acoustic images of 7 different materials using a range of frequencies.

We chose wood, cardboard, metal, plastic, cloth, and rubber for testing materials as they have different textures, and therefore, were used to examine the system's ability to distinguish the object and assess its response. We also tested the imaging system with a human's hand in order to analyze its performance for objects with sophisticated textures. Each material was imaged using a range of frequencies from 1 kHz to 7 kHz, with a 0.5 kHz interval. To avoid spatial aliasing, the maximum frequency for the transmitted signal should not be higher than 7391.3 Hz for a sensor spacing of 2.3 cm. Thus, in our system, it was set to 7000 Hz. The reflected signals from the tested object were received by the 64-node array, and then processed by the DSBF algorithm. The image was generated using the same procedure described in sound localization.

The imaged results of the 7 materials under different frequencies are shown in Fig. 4. The same wooden board with 8×8 grid were used for both testing object and calibration. We first recorded the reflected signals from this wooden board, and obtained its sound images. By comparing these images with the results of the sound localization experiment (see Fig. 3 (a)), we were able to relatively locate where the source signal were reflected from in the grid. All the other materials with smaller size were then glued over this specific area in the grid. This allowed us to calibrate the positions between the object and the sound source so that the signals received by the array were the direct reflections from the object itself.

To further analyze the acoustic responses between different materials, we have plotted the frequency response curves of these materials in Fig. 5, according to their relative signal intensities at the reflection area. For frequencies above 3 kHz, the peak value in the image is the direct response power from the object, while for frequencies between 1 kHz to 3 kHz, we need to first locate the relative reflection area in its acoustic image by utilizing the object location information we acquired during the calibration process, and then measure the signal intensity level in the middle of this area as the estimation of the object's response power. However, at lower frequencies, the wavelength of the sound source becomes comparable to the size of the object. This will cause diffraction, which makes the results relatively inaccurate. Therefore, the curves between 1 kHz

and 2 kHz are plotted in dot line because they may not necessarily represent the real response power from the object.

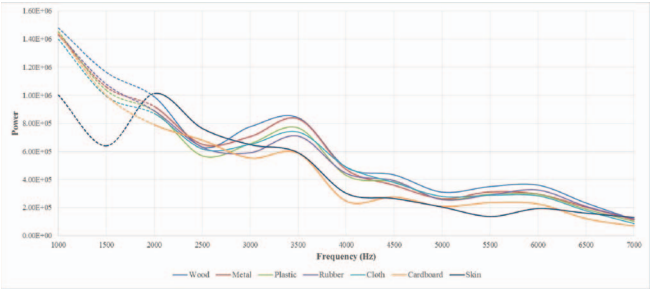

Fig. 5. Frequency response curves of seven different materials.

It can be seen that each material has its unique response with the increase of the frequency. Overall, wood, metal, plastic, cloth, and rubber have good sound reflecting properties, while the signal power from cardboard and human skin is relatively weak from 3.5 kHz to 6.5 kHz. This is likely due to the complex texture of the human skin (and cardboard), which scatters or absorbs most of the energy so that the detected echo represents only a small portion of the original signal. The result indicates that generally different textures will have different acoustic response powers. Our system can detect the difference between textures based on their unique response powers at certain frequency. More specifically, from 3.7 kHz to 4.3 kHz, and 5 kHz to 6 kHz, the response powers of human skin and cardboard present a significant difference compared to other materials. Such differences can also be observed in Fig. 4 at 4 kHz, 5 kHz and 5.5 kHz, where the sound images generated from the human skin and cardboard can be clearly distinguished from other materials.

6. CONCLUSIONS

In this paper, we have examined the design, implementation, and evaluation of a 64-node microphone array system that is capable of achieving acoustic imaging. Our array system is able to locate the sound source with an average error of 1.1 degrees. It can also separate sound sources with a mean error of 13.1 degrees. Both experiments were conducted in the presence of noise and reverberation, which proves our system's robustness for real world applications. We have presented the acoustic images of seven different materials as well as their frequency response curves from 1 kHz to 7 kHz. The reconstructed images generally show a similar visual response pattern with the increase of frequency, while a relatively distinct response can be observed from human skin and cardboard at certain frequency. Furthermore, from the frequency response curves, we are able to acquire the reflection power of different materials from 2 kHz to 7 kHz, which can then be utilized to distinguish the objects. Our results show that generally materials with smooth texture have a strong response power, and those with rough texture have a relatively weak response. These unique responses will help to build the representation of different objects in the environment, which can be useful for object detection and recognition.

For future work, we plan to collect more reflection patterns from more different textures with various sizes. Additionally, since the sound source itself (such as its power, its distance and orientation relative to the array) and also the shape of the object may affect the reflections, we hope to explore these factors and understand their effects on the acoustic responses. We also plan to integrate our array system with machine learning techniques in order to further examine its efficacy in recognizing different materials.

7. REFERENCES

[1] L. Pei, L. Chen, R. Guinness, J. Liu, H. Kuusniemi, Y. Chen, ..., and S. Soderholm, "Sound Positioning Using a Small-Scale Linear Microphone Array," In *IEEE Int. Conf. on Indoor Positioning and Indoor Navigation*, pp. 1-7, Oct. 2013.

[2] M. Goseki, H. Takemura, and H. Mizoguchi, "Visualizing Sound Pressure Distribution by Kinect and Microphone Array," In *IEEE Int. Conf. on Robotics and Biomimetics*, pp. 1243-1248, Dec. 2011.

[3] M. Goseki, M. Ding, H. Takemura, and H. Mizoguchi, "Combination of Microphone Array Processing and Camera Image Processing for Visualizing Sound Pressure Distribution," In *IEEE Int. Conf. on Systems, Man, and Cybernetics*, pp. 139-143, Oct. 2011.

[4] A. O. Donovan, R. Duraiswami, and D. Zotkin, "Imaging Concert Hall Acoustics Using Visual and Audio Cameras," In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 5284-5287, Mar. 2008.

[5] M. Legg, and S. Bradley, "A Combined Microphone and Camera Calibration Technique with Application to Acoustic Imaging," *IEEE Trans on Image Processing*, vol. 22, no. 10, pp. 4028-4039, Oct. 2013.

[6] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, 2004.

[7] M. Turqueti, J. Saniie, and E. Oruklu, "MEMS Acoustic Array Embedded in an FPGA Based Data Acquisition and Signal Processing System," In *IEEE Int. Midwest Symp. on Circuits and Systems*, pp. 1161-1164, Aug. 2010.

[8] H. Kajbaf, and H. Ghassemian, "Acoustic Imaging of Heart Using Microphone Arrays," In *Int. Conf. on Biomedical Engineering*, Springer, Berlin Heidelberg, pp. 738-741, 2009.

[9] M. Moebus, and A. M. Zoubir, "Three-Dimensional Ultrasound Imaging in Air Using a 2D Array on a Fixed Platform," In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 961-964. Apr. 2007.

[10] M. Moebus, A. M. Zoubir, and M. Viberg, "Parametrization of Acoustic Images for the Detection of Human Presence by Mobile Platforms," In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 3538-3541, Mar. 2010.

[11] I. Dokmanic, and I. Tashev, "Hardware and Algorithms for Ultrasonic Depth Imaging," In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 6702-6706, May 2014.

[12] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*, Springer Science & Business Media, 2008.

[13] B. D. Van Veen, and K. M. Buckley, "Beamforming: A Versatile Approach to Spatial Filtering," *IEEE ASSP Magazine*, pp. 4-24, 1988.

[14] D. H. Johnson, and D. E. Dudgeon, *Array Signal Processing*, Prentice Hall, New Jersey, 1993.