Proceedings of the 2020 Winter Simulation Conference
K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, eds.

COMPARISON OF DIFFERENT BEAMFORMING-BASED APPROACHES FOR SOUND SOURCE SEPARATION OF MULTIPLE HEAVY EQUIPMENT AT CONSTRUCTION JOB SITES

Behnam Sherafat
Abbas Rashidi

Department of Civil and Environmental Engineering
University of Utah
201 Presidents' Cir
Salt Lake City, UT 84112, USA

Sadegh Asgari

Department of Civil Engineering
Merrimack College
315 Turnpike St,
North Andover, Massachusetts 01845, USA

ABSTRACT

Construction equipment performance monitoring can support detecting equipment idle time, estimating equipment productivity rates, and evaluating the cycle time of activities. Each equipment generates unique sound patterns that can be used for equipment activity detection. In the last decade, several audio-based methods are introduced to automate the process of equipment activity recognition. Most of these methods only consider single-equipment scenarios. The real construction job site consists of multiple machines working simultaneously. Thus, there is an increasing demand for advanced techniques to separate different equipment sound sources and evaluate each equipment's productivity separately. In this study, six beamforming-based approaches for construction equipment sound source separation are implemented and evaluated using real construction job site data. The results show that Frost beamformer and time-delay Linear Constraint Minimum Variance (LCMV) generate outputs with array gains of more than 4.0, which are more reliable than the other four beamforming techniques for equipment sound separation.

1 INTRODUCTION

The construction industry is underperforming relative to other industries and the whole economy; Construction productivity, in most advanced economies, has been flat while manufacturing productivity has almost doubled over the past two decades (Changali et al. 2015). To achieve higher construction productivity, project managers need to improve on-site execution through tracking the efficient utilization of labor, material, and equipment (Barbosa et al. 2017). Researches have been working on increasing the productivity rate of the on-site construction activities by introducing automated approaches in different construction domains (Noghabaei et al. 2020; Zamen and Dehghan-Niri 2020).

Heavy equipment is a critical and costly resource for contractors. In the last decade, the number of research studies on construction equipment performance monitoring has increased considerably. Three different approaches have been introduced: 1) vision-based methods (Zhu et al. 2017; Roberts and Golparvar-Fard 2019; Luo et al. 2020), 2) audio-based methods (Sherafat et al. 2019a, b; Chen et al. 2020; Lee et al. 2020), and 3) kinematic-based methods (Rashid and Louis 2019; Rashid and Louis 2020; Slaton et al. 2020). Among these approaches, recognizing construction activities using generated sound patterns have shown significant potential and proved to be efficient and economical because their required devices (e.g., microphone and microphone arrays) are inexpensive and accessible (Sabillon et al. 2020). Moreover, these methods can be used in different job site conditions (e.g., night time) and in large construction job sites with the presence of obstacles (Sherafat et al. 2020). The major limitation of these methods, however, is that they have been tested on single-equipment scenarios without any interference or noise from other machines or workers on the job site, whereas real construction job sites include different types and numbers of machines working simultaneously, which increases the need for more robust and comprehensive methods.

In real construction job sites, several equipment sounds are being generated from different directions and distances. A microphone is installed in the construction job site to record equipment sounds. Currently, two types of microphones are available in the market: 1) single-channel microphones and 2) microphone arrays. The aforementioned audio-based methods are mostly using single-channel microphones. Single-channel microphones are emphasizing on signals coming from a specific direction while attenuating signals coming from other directions. These types of microphones have a fixed directivity pattern, which should be physically directed towards the desired equipment sound signal (Allred 2006). These microphones lack spatial information about sounds in the environment and perform poorly when several sound sources are available (Schwartz et al. 2017). In other words, these types of microphones cannot find the Direction of Arrival (DOA) and the location of the sound sources, thus making them unsuitable for multiple-equipment scenarios. On the other hand, microphone arrays are capable of separating specific sounds and enhancing the Signal to Noise Ratio (SNR) using acoustic beamforming. Based on the type of microphone being used, sound source separation methods are classified into two main categories: 1) software-based methods and 2) hardware-based methods. Software-based methods use single-channel microphones and need special mathematical algorithms, whereas hardware-based methods use microphone arrays to detect the DOA of signals using beamforming techniques.

Acoustic beamforming is one of the main methods of spatial filtering or localization of the desired sound source from a variety of other unwanted sound sources in the environment based on DOA using microphone arrays. Beamforming algorithms work based on relative time delays between different microphones. It has a wide variety of applications, including speech recognition (Schwartz et al. 2017; Holder et al. 2020), hearing aids (Doclo et al. 2008), and robot systems (Novoa et al. 2019).

In the past decades, beamforming techniques have been applied in different areas. Speech enhancement methods using microphone arrays and beamforming techniques have been studied extensively (Dam and Nordholm 2017; Xenaki et al. 2018). Tiana-Roig et al. (2010) used the classical delay-and-sum beamforming and circular harmonics beamforming to localize different environmental noise sources. Dam et al. (2016) used an optimal beamformer to extract the desired speech signal in a two-speaker environment; They used a voice activity detector to find the activity information and location of the speakers to feed their optimal beamformer. In the near-field research domain, Kumar and Hedge (2016) investigated different methods of beamforming for near-field acoustic source localization. Research studies in different application areas are trying to improve the performance and optimize the algorithm for real-time purposes.

Before using beamforming techniques in these applications, different types of conditions should be considered, which might affect the results of the beamforming. One of the crucial factors is the distance between the microphone array and sound sources. Sound sources that are farther than specific criteria are considered as far-field, while sound source which is closer are considered as near-field. The former has a spherical wavefront, while the latter has a planar wavefront. The other significant factor is the location of the sound sources. Sound sources which are generated indoor required specific considerations due to the sound reverberations. Beamforming methods can be classified based on different properties which are as follows:

1) Conventional or adaptive: Conventional beamforming or classical beamforming techniques do not have fixed weights for beam pattern and do not depend on the array data (e.g., time delay beamformer). On the other hand, adaptive beamforming techniques find weights based on the array data (e.g., frost beamformer).

2) Narrowband or wideband: Bandwidth of the signal affects choosing the appropriate type of beamforming technique. Narrowband beamformers (e.g., phase shift beamformer) are treated differently than wideband beamformers (e.g., generalized side-lobe canceler beamformer) because they assume that signals contain much less frequency spectrum. When the signal is narrowband, time delay in time-domain is equivalent to a phase shift in frequency-domain.

3) Time-domain or frequency-domain: Signals can be processed in time-domain, or they can be converted to the frequency domain, a linear phase shift is applied, and then they converted back to the time-domain.

In the construction domain, to the best of our knowledge, sound localization and separation have not been thoroughly studied. In a recent study, Cheng et al. (2019) used delay-and-sum beamforming to isolate the desired signal from a mixed-signal. They have not, however, provided a detailed analysis of the assumptions and the results of the beamforming. This study aims to investigate the results of different beamformers on the construction job sites to separate different equipment sound sources from the mixed signals recorded using a microphone array.

The rest of this study is organized as follows: Section 2 explains several beamformers. Section 3 discusses the implementation of these beamformers on real construction job sites with multiple equipment working simultaneously, and Section 4 presents the results. Finally, Section 5 provides the conclusion and suggestions for future studies.

2 BEAMFORMERS

Based on the introduction to beamformers, the authors evaluated the equipment signals on the construction job site to have a better understanding of the signal in time-domain and frequency-domain. The recorded signals of three types of machines (i.e., bulldozer, jackhammer, and excavator) in both time and frequency-domain are shown in Figure 1 and Figure 2, respectively. $|P_1(f)|$ shows the magnitude of each frequency in the frequency-domain. These signals are recorded at 44.1 kHz for 30 seconds separately without any interference from other machines.

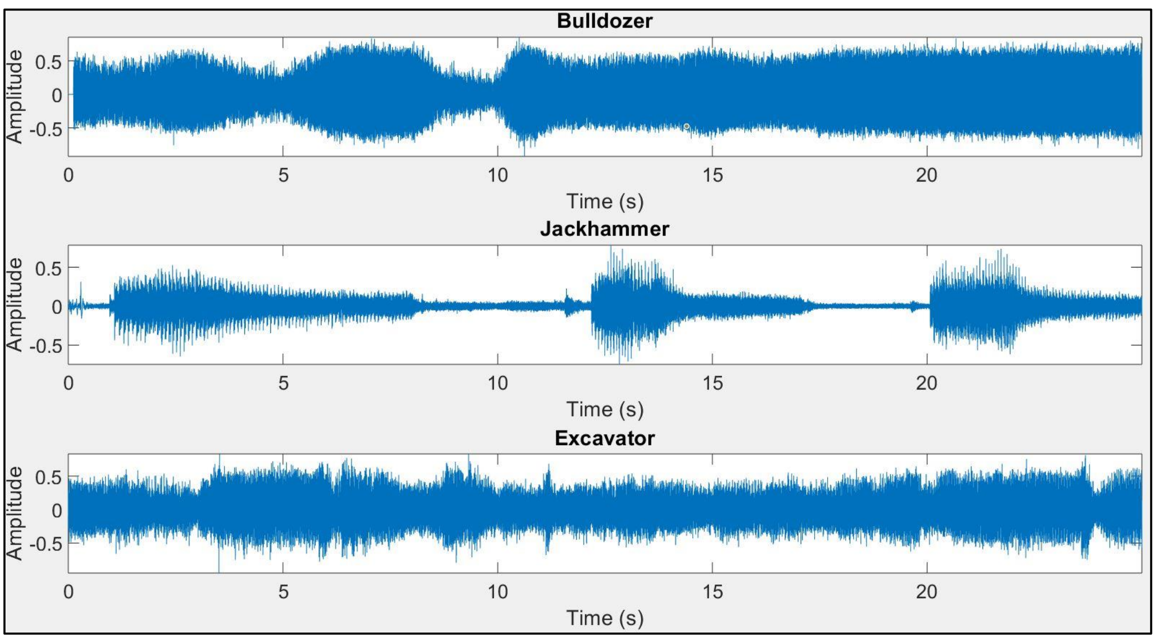

Figure 1: Time-domain signal of three types of construction equipment.

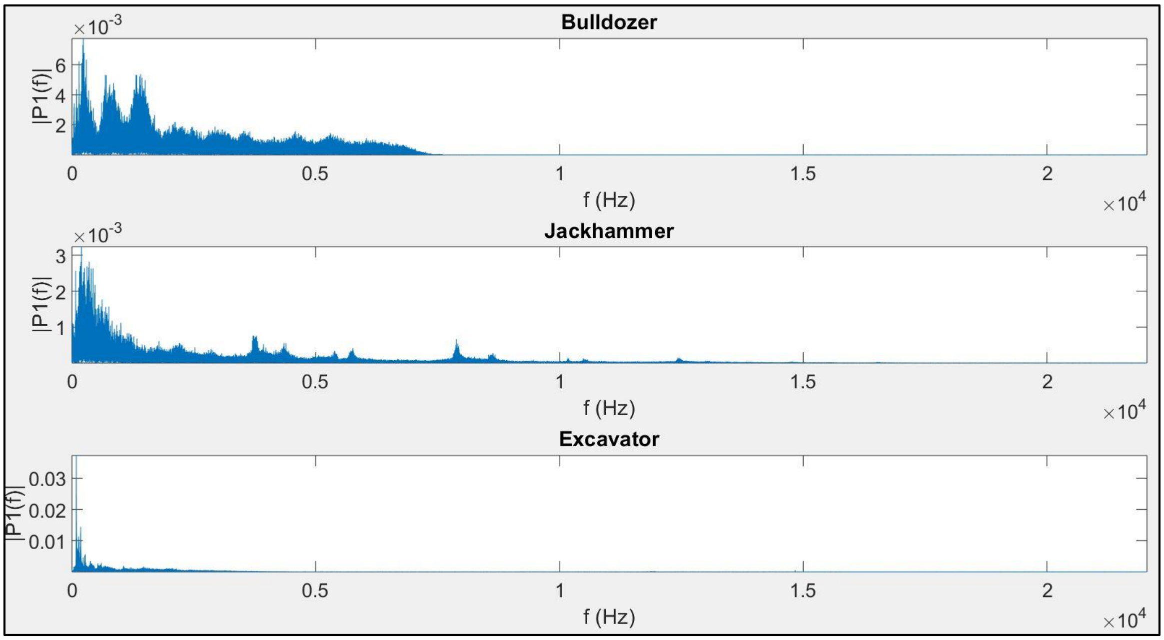

Figure 2: Single-sided amplitude spectrum of three types of construction equipment signal.

A careful examination of Figure 2 shows that different construction equipment generates different types of signals in both time-domain and frequency-domain. For example, bulldozer, jackhammer, and excavator signals have negligible frequency components above 7 kHz, 17 kHz, and 3kHz, respectively. However, construction equipment signals still consist of a wide range of frequency components, which confirms that their signals are wideband. Thus, to separate equipment sound sources in the construction job site, beamformers that are specific for wideband signals should be chosen. This study compares six common wideband beamformers, which are as follows: 1) time-delay beamformer, 2) sub-band phased shift beamformer, 3) time-delay Linear Constraint Minimum Variance (LCMV) beamformer 4) frost beamformer, 5) generalized side-lobe canceler (GSC) beamformer, and 6) wideband Minimum-Variance Distortionless-Response (MVDR) beamformer. In the following sections, a high-level structure of these beamformers is introduced.

2.1 Time-delay Beamformer

The Time-delay beamformer is the simplest type of beamformer. It performs delay-and-sum beamforming based on the relative delays between the plane-wave signals arriving at the array microphones. Microphones are receiving the same signal from a specific direction but with different delays. This beamformer calculates the delays and applied reverse delays to the signal. In Figure 3, the architecture of the delay-and-sum beamforming is shown. The extra distance that the signal needs to propagate, Δ, is calculated using equation (1).

$$\Delta = a \sin \vartheta \tag{1}$$

In equation (1), a is the distance between microphones in the microphone array in meters, and ϑ is the angle of incidence of the signal in degree. This distance can be expressed as radian measure, φ, using equation (2).

$$\varphi = \frac{2\pi}{\lambda} a \sin \vartheta \tag{2}$$

In equation (2), λ is the wavelength of the signal. This phase angle is used to delay or advance the signal using equation (3).

$$e^{i\varphi}e^{i\omega t} = e^{i(\omega t+\varphi)} \tag{3}$$

In equation (3), ω is the frequency of the signal, and t is the time. The signals received by the microphones are delayed or advanced in the time-domain or frequency-domain, accordingly, summed together, divided by the appropriate weight coefficients, and converted back to the time-domain if these calculations are done in the frequency-domain. In delay-and-sum beamforming, all of the weight coefficients for microphones are the same and equal to $1/n$ in which n is the number of microphones.

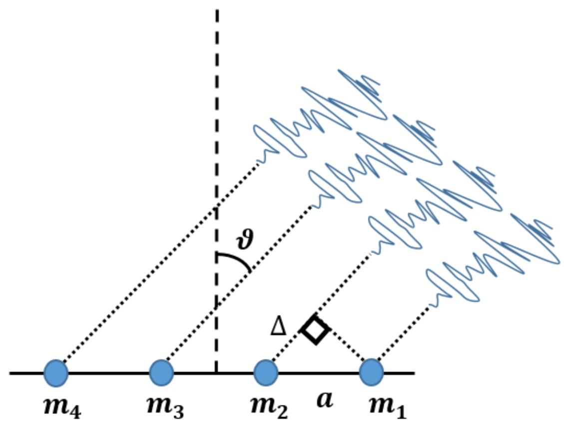

Figure 3: Delay-and-sum beamforming architecture.

2.2 Sub-band Phased Shift Beamformer

This beamformer splits the wideband signal into sub-bands and applies narrowband phased shift beamforming in each of the sub-bands (Kumatani et al. 2008). Using the same process as delay-and-sum beamformer, the delays in the time-domain are converted to phased shifts in the frequency-domain, and each microphone signal is multiplied by a frequency-dependent phase shift that compensates for the delay. After all of the signals in the sub-bands are delayed or advanced, they are rearranged together to generate the output signal.

2.3 Time-delay Linear Constraint Minimum Variance (LCMV) Beamformer

This beamformer implements the LCMV beamformer in the time-domain by steering the array to the desired direction and applying a Finite Impulse Response (FIR) filter to the output of each microphone. The LCMV beamformer calculates the weights by minimizing the total output power of the array, P, and considering some constraints (Van Trees 2002) using equation (4).

$$P = w^H S w \tag{4}$$

In equation (4), S, is the sensor spatial correlation matrix, H is the Hermitian matrix or conjugate transpose of it, and w is the weight matrix. A constraint is applied to the weights to generate a unit gain in the direction of the desired signal, and some constraints may be applied to other directions to nullify the array response in those directions. $(Az_1, El_1), (Az_2, El_2), \ldots (Az_K, El_K)$ are all known directions with a constraint. Each direction has its own corresponding steering vector, c_k, and the response of the array to that direction, r_k is calculated using equation (5).

$$c_k^H w = r_k \tag{5}$$

In equation (4), H is the transpose conjugate. Concatenating all of these constraints into matrix C and responses into a single matrix R, the matrix form of equation (4) is derived, which is shown in equation (6).

$$C^H w = R \tag{6}$$

LCMV minimizes equation (4) subject to the above constraints to find the weights. The weights are calculated using equation (7).

$$w = S^{-1}C(C^H S^{-1} C)^{-1} R \tag{7}$$

2.4 Frost Beamformer

Frost algorithm (Frost 1972) has been recognized as one of the popular adaptive beamformers as it is capable of reduces the noise from other directions while preserving the desired signal from the incident direction. It adaptively minimizes the noise energy by finding appropriate weights for the FIR filter for the received and delayed signal at each microphone in the array. This beamformer only needs the desired DOA and frequency band of the signal. In other words, Frost algorithm is a type of LCMV beamformer that utilizes the stochastic gradient descent algorithm to iteratively find the statistics of the noise arriving from different angels (Ribeiro et al. 2019). The LCMV beamformer maintains the output power of the microphone array in the desired direction while suppressing the power from other directions. In other words, Frost algorithm The structure of the Frost beamformer is shown in Figure 4.

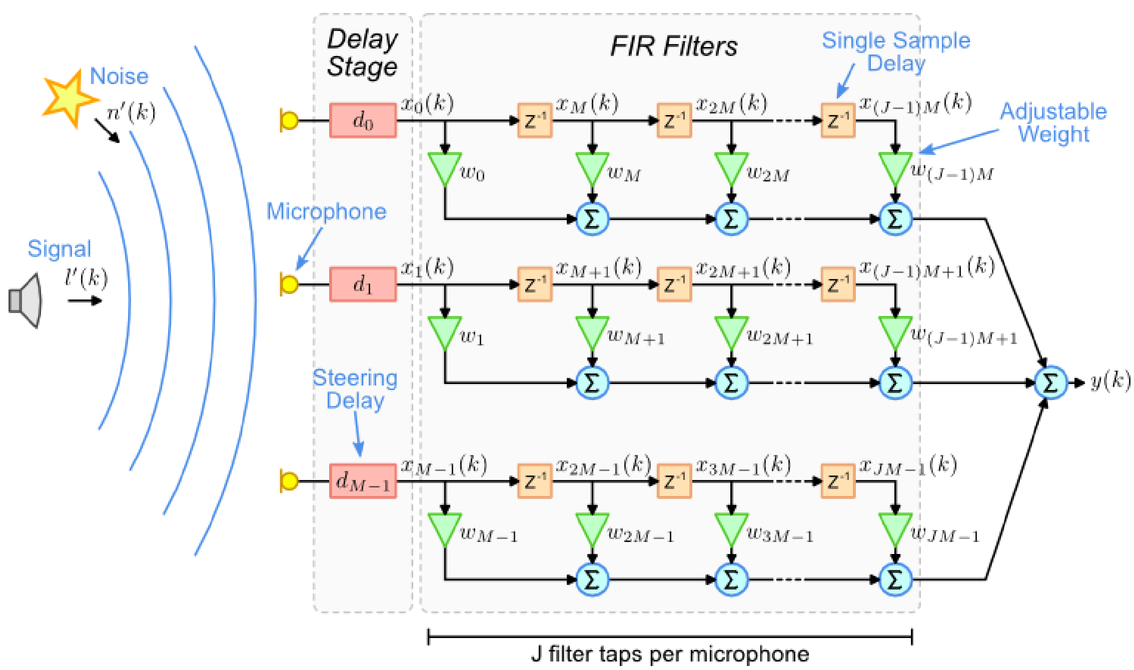

Figure 4: The filter-sum beamformer structure for Frost's algorithm (Greensted 2010).

This beamformer consists of two stages: 1) the delay stage and 2) the FIR stage. In the delay stage, the signals received by the microphones are aligned. In the FIR filter stage, weights are adjusted to maintain the input signal power. X^T, the transpose of the sample vector and W^T, the weight vector are shown in equation (8) and equation (9).

$$X^T(k) \triangleq [x_0(k), x_1(k), \dots, x_{M-1}(k), x_M(k), \dots, x_{2M-1}(k), \dots, x_{(J-1)M}(k), \dots, x_{JM-1}(k)] \tag{8}$$

$$W^T(k) \triangleq [\omega_0, \omega_1, \dots, \omega_{M-1}, \omega_M, \dots, \omega_{2M-1}, \dots, \omega_{(J-1)M}, \dots, \omega_{JM-1}] \tag{9}$$

The output of the beamformer $y(k)$ is the multiplication of sample and weight vectors using equation (10). Also, a constraint matrix is used for the weights in a way that only minimizes the noise power, not the desired signal power.

$$y(k) = [\omega_0, \dots, \omega_{JM-1}][x_0(k), \dots, x_{JM-1}(k)]^T \tag{10}$$

2.5 Generalized Side-lobe Canceler (GSC) Beamformer

This type of beamformer implements the LCMV beamformer efficiently. When the number of microphones increases, the LCMV beamformer is computationally costly. GSC beamformer utilizes an alternative approach by converting the problem to an adaptive unconstraint problem (Griffiths and Jim 1982). The GSC beamformer splits incoming signals into two signal paths. The upper path is a Fixed Beamformer (FB), which applies a conventional beamformer while the lower path implements an adaptive unconstrained beamformer. The Blocking Matrix (BM) is orthogonal to the signal and removes the signal and maintains the noise from other directions. The output of the GSC beamformer is calculated by subtracting the output of the lower path from the output of the upper path. The structure of the GSC beamformer is shown in Figure 5.

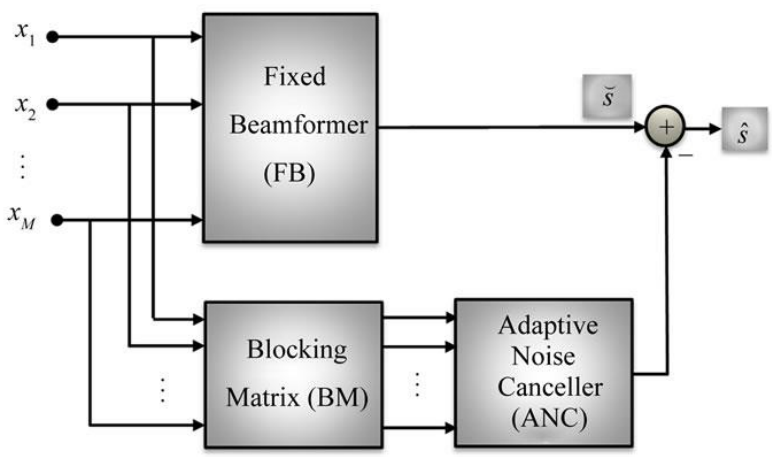

Figure 5: Structure of the GSC beamformer (Hidri et al. 2012).

2.6 Wide-band Minimum-Variance Distortionless-Response (MVDR) Beamformer

MVDR beamformer is a data-dependent adaptive beamformer to minimize the variance of the recorded signal. Because usually the desired signal and the noise are uncorrelated, the variance of the recorded signal is the sum of the variances of the noise and the desired signal (Van Trees 2002). MVDR tries to minimize this sum and suppress the noise signals. MVDR weights, w, are calculated using equation (11) in a recursive way.

$$w = \frac{S^{-1}V_0}{{V_0}^H S^{-1} V_0} \tag{10}$$

In equation (10), V_0 is the steering vector corresponding to the desired direction, and S is the spatial covariance matrix.

3 METHODOLOGY

To separate different equipment sound sources on the construction job site, a microphone array is required. As mentioned before, single microphones are not appropriate for sound source separations, and most of the current studies on sound source separation are implemented using microphone arrays. Microphone arrays consist of several microphones with specific layouts such as linear, rectangular, circular, and conformal. In this study, the authors used XMOS xCORE-200 (XMOS, Bristol, England) and placed it on the job site with distances less than 15 m from the sound sources. This microphone array consists of seven microphones, six on the perimeter and one in the middle (Figure 6). The authors have only used the recordings of six microphones on the perimeter. Each microphone receives the same signal with different delays and levels of noise. Also, the quality of the recorded sound is dependent on the distance of the sound sources and the presence of noise in the environment.

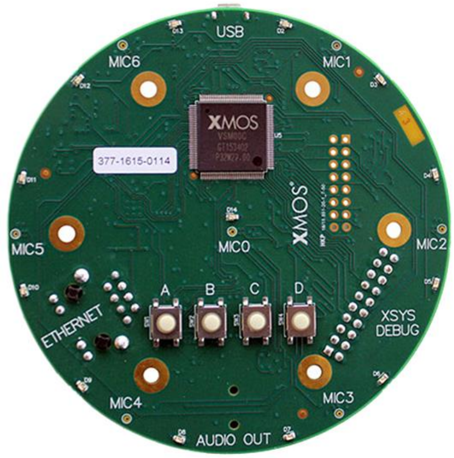

Figure 6: XMOS xCORE-200 microphone array.

In this study, 25 seconds of recordings are used to apply different beamformers. The sound speed is used for calculations, and the sampling frequency is 44.1 kHz. Also, to separate sound sources, a time window of 2 seconds is used. This time window is enough to capture the delays between different channels, and it has sufficient frequency content for signal processing step in each beamformer. Equipment sounds are received from different directions. The sound direction angle consists of two parameters; [azimuth, elevation]. Generally, the azimuth angle is positive when it is rotated counter-clockwise from x-direction, and the angles lie between -180 degrees to 180 degrees. The elevation angle is measured between the XY plane and the vector pointing the object and lies between -90 degrees and 90 degrees. To evaluate the performance of beamformers, array gain has been used as the output criteria. The sounds of each machine are also recorded separately without any interference from other machines to be used in array gain calculation. Array gain is the ratio of output Signal-to-Interference-plus-Noise Ratio (SINR) to input SINR and is measured in decibels using equation (11) (Frost 1972). The more value for array gain means that the beamformer has better performance. For example, an array gain of 10 means that the beamformer obtained an SINR improvement of 10 dB.

$$Array\ Gain = 10\log_{10}\left(\frac{mean((Jackhammer\ input\ time\ signal + Compactor\ input\ time\ signal)^2)}{mean((Bulldozer\ output\ time\ signal - Bulldozer\ input\ time\ signal)^2)}\right) \tag{11}$$

In the next section, the results of sound sources separation for a real case study on the construction job sites are described.

4 CASE STUDY AND RESULTS

This study evaluates the performance of these beamformers on construction equipment sound separation. Three types of machines (i.e., bulldozer, jackhammer, and compactor) are working simultaneously, and a circular microphone array with six microphones is placed in the middle of them. Figure 7 presents the layout of the machines' sounds directions and microphone array setup.

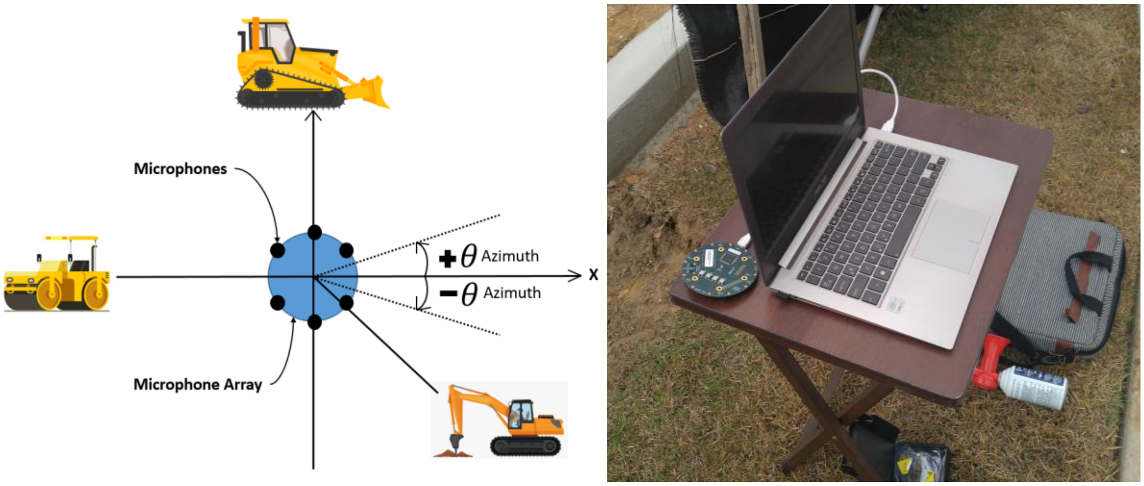

Figure 7: Layout of a machines' sounds directions (left); microphone array setup (right).

For this case study, it is assumed that the elevation angle of the sound received from the machines is zero because the machines are located far away from the microphone array. Also, due to this distance, it is assumed that sound signals are far-field, and they have a planar wavefront. The corresponding DOAs for the machines' sounds are as follows: 1) jackhammer [-45, 0]; 2) bulldozer [90, 0]; and 3) compactor [180, 0]. In this case study, the authors separate the bulldozer sound received from [90, 0] and evaluated the enhancement performance of the beamformers by their array gain.

Table 1 presents the array gain for different beamformers. Sub-band phased shift, and wideband MVDR beamformers output depends on the number of sub-bands and time-delay LCMV, Frost, and GSC beamformers rely on the filter length.

Table 1: Comparison of different beamformers.

Beamformer	Number of Sub-bands	Array gain (dB)	Beamformer	Filter Length	Array gain (dB)
Time-delay	NA	3.6680	Frost	20	4.1505
Sub-band Phased Shift	25	2.4801		40	4.1827
	50	3.1325		60	3.8350
	100	3.4679		80	3.6492
	150	3.5694		100	3.5833
	200	3.6345	GSC	20	1.6021
	250	3.6532		40	2.1882
	300	3.6838		60	2.2323
Wide-band MVDR	5	0.1831		80	2.1257
	10	0.8149		100	2.1457
	15	1.1939	Time-delay LCMV	20	4.1505
	20	1.5048		40	4.1827
	25	1.3493		60	3.8350
	30	1.0678		80	3.6492
	35	0.9103		100	3.5833

As Table 1 shows wideband MVDR has the lowest average array gain, and Frost and time-delay LCMV have the highest average array gain. Also, the array gain for Frost and time-delay beamformers are similar for the same filter length. Moreover, the performance of the sub-band phased shift beamformer increases as the number of sub-bands increases. However, this increase is not significant after 200 sub-bands. Increasing the number of sub-bands impact on the computational time of the beamformer, so it is not efficient. Wide-band MVDR has the highest array gain for 20 sub-bands, Frost and time-delay LCMV has the best performance with a filter length of 40, and GSC has its highest array gain with a filter length of 60. To better analyze the effect of the top two beamformers (i.e., sub-band phased shift, Frost), the effect of the number of sub-bands on sub-band phased shift beamformer and the effect of the filter length on are investigated. Figure 8 presents the results of this analysis.

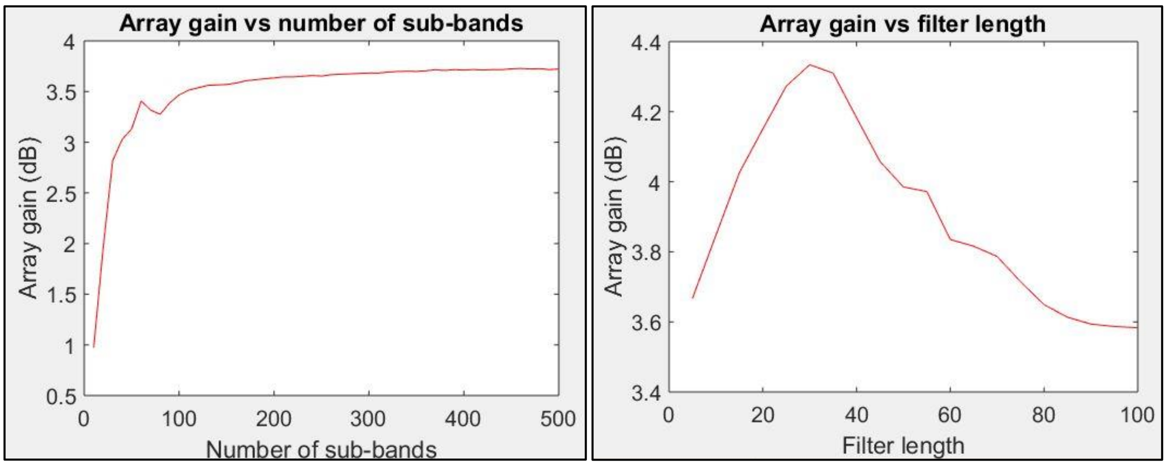

Figure 8: Array gain vs. the number of sub-bands for sub-band phased-shift beamformer (left); Array gain vs. filter length for Frost beamformer (right).

As the number of sub-bands increases, the array gain of sub-band phased-shift beamformer increases; however, this increase is not significant for the number of sub-bands more than 200. Also, the array gain for Frost has the highest value for filter length of 30, and it decreases by increasing the filter length.

5 CONCLUSIONS

Beamformers have abundant applications in radar and sonar. They have not been used in the construction industry, yet. One of their applications in the construction industry is separating different sound sources on the construction job site and using separated sounds for automated activity recognition and tracking. This study utilizes different beamformers for a simple case study where three types of machines are working simultaneously. The machines' sounds are received from different angles, and the sound of bulldozer is separated using different types of beamformers. These beamformers have different performances based on different conditions such as far-field or near-field conditions, noisy environment, and also different beamformer setups such as the number of sub-bands and filter length. Several assumptions are used in this case study. For example, the signals generated by the machines are assumed to be far-field, and the elevation for DOAs is assumed to be zero due to the distance between the microphone array and the machine.

There are some limitations to this study. The number of machines and their DOAs is known before running beamformers. In other words, these two parameters are the inputs of the beamformers and should be known or estimated. Also, this study assumes that multiple machines are working in different sound directions. Separating sound sources coming from the same direction needs more complicated setup, such as placing several microphones on the job site and recording the equipment sounds from multiple directions.

There are several methods available for DOA estimation, which will be evaluated in future studies. Construction job sites are noisy environments that impact on the performance of beamforming. Furthermore, some construction job sites are large and require several microphone arrays to be placed to record different sound sources from different corners of the job sites because sound power decreases as it propagates through space.

ACKNOWLEDGMENTS

This research project has been funded by the U.S. National Science Foundation (NSF) under Grant CMMI-1606034. The authors gratefully acknowledge NSF's support. Any opinions, findings, conclusions, and recommendations expressed in this manuscript are those of the authors and do not reflect the views of the funding agency. The authors also appreciate the assistance of Jeremy Jareckyj and Cami Richman, undergraduate students at the University of Utah, with data collection and audio recordings.

REFERENCES

Allred, D. J. 2006. *Evaluation and Comparison of Beamforming Algorithms for Microphone Array Speech Processing*. Ph.D. thesis, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia, https://smartech.gatech.edu/handle/1853/11606, accessed 15th February 2020

Barbosa, F., J. Woetzel, J. Mischke, M. J. Ribeirinho, M. Sridhar, M. Parsons, N. Bertram, and S. Brown. 2017. *Reinventing Construction: a Route to Higher Productivity*. McKinsey & Company, McKinsey Global Institute, New York, New York, https://www.mckinsey.com/industries/capital-projects-and-infrastructure/our-insights/reinventing-construction-through-a-productivity-revolution. accessed 20th February 2020

Ceolini, E., J. Anumula, A. E. Huber, I. Kiselev, and S. C. Liu. 2018. "Speaker Activity Detection and Minimum Variance Beamforming for Source Separation." In *Interspeech*, September 2nd-6th, Hyderabad, India, 836-840.

Changali, S., A. Mohammad, and M. V. Nieuwland. 2015. *The Construction Productivity Imperative*. *How to Build Megaprojects* Better. McKinsey & Company, McKinsey Productivity Sciences Center, New York, New York. https://www.mckinsey.com/industries/capital-projects-and-infrastructure/our-insights/the-construction-productivity-imperative, accessed 21th February 2020

Chen, C., Z. Zhu, and A. Hammad. 2020. "Automated Excavators Activity Recognition and Productivity Analysis from Construction Site Surveillance Videos". *Automation in Construction* 110, p.103045.

Cheng, C. F., A. Rashidi, M. A. Davenport, and D. V. Anderson. 2019. "Evaluation of Software and Hardware Settings for Audio-Based Analysis of Construction Operations". *International Journal of Civil Engineering* 17(9):1469-1480.

Dam, H. Q .H. and S. Nordholm. 2017. "Source Separation Employing Beamforming and SRP-PHAT Localization in Three-Speaker Room Environments". *Vietnam Journal of Computer Science* 4(3):161-170.

Dam, H. Q. H., H. Ho, and M. H. Le Ngo. 2016. "Blind Speech Separation Using SRP-PHAT Localization and Optimal Beamformer in Two-Speaker Environments." *International Journal of Computer and Information Engineering* 10(8):1529-1533.

Doclo, S., S. Gannot, M. Moonen, A. Spriet, S. Haykin, and K. R. Liu. 2008. "Acoustic beamforming for hearing aid applications." In *Handbook on Array Processing and Sensor Networks*, edited by S. Haykin and K. J. R. Liu, 269-302. New York: Wiley- Institute of Electrical and Electronics Engineers Press

Frost, O. L. 1972. "An Algorithm for Linearly Constrained Adaptive Array Processing". *Journal of Proceedings of the Institute of Electrical and Electronics Engineers* 60(8):926-935.

Greensted, A. 2010. "Frost's Algorithm", http://www.labbookpages.co.uk/audio/beamforming/frost.html, accessed 15th February

Griffiths, L. and C. W. Jim. 1982. "An Alternative Approach to Linearly Constrained Adaptive Beamforming". *Institute of Electrical and Electronics Engineers: Transactions on Antennas and Propagation* 30(1):27-34.

Hidri, A., S. Meddeb, and H. Amiri. 2012. "About Multichannel Speech Signal Extraction and Separation Techniques". *Journal of Signal and Information Processin* 3, *pp.238-247.*

Holder, J. T., A. L. Taylor, L. W. Sunderhaus, and R. H. Gifford. 2020. "Effect of Microphone Location and Beamforming Technology on Speech Recognition in Pediatric Cochlear Implant Recipients". *Journal of the American Academy of Audiology* 31(7):506-512

Kumar, L. and R. M. Hegde. 2016. "Near-Field Acoustic Source Localization and Beamforming in Spherical Harmonics Domain". *Institute of Electrical and Electronics Engineers: Transactions on Signal Processing* 64(13):3351-3361.

Kumatani, K., J. McDonough, S. Schacht, D. Klakow, P. N. Garner, and W. Li. 2008. "Filter Bank Design Based on Minimization of Individual Aliasing Terms for Minimum Mutual Information Subband Adaptive Beamforming." In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, March 31st-April 4th, Las Vegas, Nevada, USA, 1609-1612.

Lee, Y. C., M. Shariatfar, A. Rashidi, and H. W. Lee. 2020. "Evidence-Driven Sound Detection for Prenotification and Identification of Construction Safety Hazards and Accidents". *Automation in Construction* 113, p.103127.

Luo, H., M. Wang, P. K. Y. Wong, and J. C. Cheng. 2020. "Full Body Pose Estimation of Construction Equipment Using Computer Vision and Deep Learning Techniques." *Automation in Construction* 110, p.103016.

Novoa, J., R. Mahu, A. Díaz, J. Wuth, R. Stern, and N. B. Yoma. 2019. "Weighted delay-and-sum beamforming guided by visual tracking for human-robot interaction". https://arxiv.org/abs/1906.07298, accessed 13th February.

Noghabaei, M., A. Heydarian, V. Balali, and K. Han. 2020. "Trend Analysis on Adoption of Virtual and Augmented Reality in the Architecture, Engineering, and Construction Industry". *Data* 5(1):26.

Rashid, K. M. and J. Louis. 2019. "Times-Series Data Augmentation and Deep Learning for Construction Equipment Activity Recognition". *Advanced Engineering Informatics* 42, p.100944.

Rashid, K. M. and J. Louis. 2020. "Automated Activity Identification for Construction Equipment Using Motion Data From Articulated Members. Front". *Built Environ* 5, p.144.

Ribeiro, L. N., A. L. de Almeida, and J. C. M. Mota. 2019. "Separable Linearly Constrained Minimum Variance Beamformers". *Signal Processing* 158, pp.15-25.

Roberts, D. and M. Golparvar-Fard. 2019. "End-To-End Vision-Based Detection, Tracking and Activity Analysis of Earthmoving Equipment Filmed at Ground Level." *Automation in Construction* 105, p.102811.

Sabillon, C., A. Rashidi, B. Samanta, M. A. Davenport, and D. V. Anderson. 2020. "Audio-Based Bayesian Model for Productivity Estimation of Cyclic Construction Activities." *Journal of Computing in Civil Engineering* 34(1):04019048.

Schwartz, O., S. Gannot, and E. A. Habets. 2017. "Multispeaker LCMV Beamformer and Postfilter for Source Separation and Noise Reduction." *Institute of Electrical and Electronic Engineers/Association for Computing Machinery Transactions on Audio, Speech, and Language Processing* 25(5):940-951.

Sherafat, B., A. Rashidi, Y. C. Lee, and C. R. Ahn. 2019a. "A Hybrid Kinematic-Acoustic System for Automated Activity Detection of Construction Equipment." *Sensors*, *19*(19), p.4286.

Sherafat, B., A. Rashidi, Y. C. Lee, and C. R. Ahn. 2019b. "Automated Activity Recognition of Construction Equipment Using a Data Fusion Approach". In *2019 Proceedings of the Computing in Civil Engineering: Data, Sensing, and Analytics*, edited by Y. K. Cho, F. Leite, A. Behzadan, C. Wang, 1-8. Reston, Virginia: American Society of Civil Engineers.

Sherafat, B., C. R. Ahn, R. Akhavian, A. H. Behzadan, M. Golparvar-Fard, H. Kim, Y. C. Lee, A. Rashidi, and E. Rezazadeh Azar. 2020. "Automated Methods for Activity Recognition of Construction Workers and Equipment: State-of-the-Art Review". *Journal of Construction Engineering and Management, 146*(6).

Slaton, T., C. Hernandez, and R. Akhavian. 2020. "Construction Activity Recognition with Convolutional Recurrent Networks". *Automation in Construction* 113, p.103138.

Tiana-Roig, E., F. Jacobsen, and E. F. Grande. 2010. "Beamforming with a Circular Microphone Array for Localization of Environmental Noise Sources." *The Journal of the Acoustical Society of America* 128(6), pp.3535-3542.

Van Trees, H. L. 2004. *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. Hoboken, New Jersey: John Wiley & Sons.

Xenaki, A., J. Bünsow Boldt, and M. Græsbøll Christensen. 2018. "Sound Source Localization and Speech Enhancement with Sparse Bayesian Learning Beamforming". *The Journal of the Acoustical Society of America* 143(6):3912-3921.

Zamen, S. and E. Dehghan-Niri. 2020. "Fractal Analysis of Nonlinear Ultrasonic Waves in Phase-Space Domain as a Quantitative Method for Damage Assessment of Concrete Structures". *Nondestructive Testing and Evaluation International* 111, p.102235.

Zhu, Z., X. Ren, and Z. Chen. 2017. "Integrated Detection and Tracking of Workforce and Equipment from Construction Jobsite Videos". *Automation in Construction* 81, pp.161-171.

AUTHOR BIOGRAPHIES

BEHNAM SHERAFAT is a Ph.D. student in Construction Engineering program at the Department of Civil and Environmental Engineering of the University of Utah. He received his B.Sc. degree in Civil Engineering from Sharif University of Technology and M.Sc. degree in Construction Engineering and Management from University of Tehran. He has done researches about Building Information Modeling (BIM), Construction Simulation, Construction Automated Quantity Take-off, Leadership in Energy and Environmental Design (LEED). He is currently focusing on automating construction performance monitoring using equipment sound data by utilizing advanced Sound Signal Processing, Machine Learning, and Artificial Intelligence algorithms. His email address is behnam.sherafat@utah.edu.

ABBAS RASHIDI is an assistant professor in Construction Engineering program at the Department of Civil and Environmental Engineering of the University of Utah. He earned his Ph.D. in the School of Civil and Environmental Engineering at Georgia Institute of Technology. His research interests include information and sensing technologies, audio signal processing for Modeling and Analysis of Civil Infrastructure Systems, video/image processing for 3D reconstruction of civil infrastructure systems, and acoustical modeling of construction jobsites. His e-mail address is abbas.rashidi@utah.edu. His website is https://faculty.utah.edu/u6013686-Abbas_Rashidi/hm/index.hml.

SADEGH ASGARI is currently an assistant professor of Civil Engineering at Merrimack College and an Adjunct Assistant Professor at Columbia University. His research focuses on developing new technologies and decision-making tools to advance analysis and design of complex systems. He is the recipient of the Zampell family faculty fellowship and Edward G. Roddy outstanding teacher of the year award. He holds a doctoral degree in Civil Engineering from Columbia University. His e-mail address is asgaris@merrimack.edu. His website is https://www.merrimack.edu/live/profiles/766-sadegh-asgari/_p/f.