

SOFTWARE AND HARDWARE REQUIREMENTS FOR AUDIO-BASED ANALYSIS OF CONSTRUCTION OPERATIONS

Chieh-Feng Cheng¹, Abbas Rashidi², Mark A. Davenport³, David V. Anderson⁴, and Chris Sabillon⁵

¹PhD student, Georgia Institute of Technology, Atlanta, GA

²Assistant Professor, University of Utah, Salt Lake City, UT; Email: abbas.rashidi@utah.edu

³Associate Professor, Georgia Institute of Technology, Atlanta, GA

⁴Professor, Georgia Institute of Technology, Atlanta, GA

⁵Graduate Student, Georgia Southern University, Statesboro, GA

ABSTRACT

Various activities of construction equipment are associated with certain distinctive sound patterns (e.g. excavating soil, breaking rocks, sawing and drilling concrete, etc.). Considering this fact, it is possible to extract information about construction operations by recording the audio at a construction site and then processing this data to determine what activities are being performed. Audio-based analysis of construction operations requires specific hardware and software choices to achieve satisfactory performance. This paper presents results from studies of the impact of these choices on the ultimate performance on the task of interest. To begin this project, an audio-based system has been developed to recognize the routine sounds of construction machinery. This system combines a denoising algorithm to enhance audio quality and a Short-Time Fourier Transform (STFT) and Support Vector Machine (SVM) to classify the performed activities of the machines. The next step in the project evaluates three types of microphones (off-the-shelf, contact, and a multichannel microphone array) and two ways to install the microphones (placed in machines

cabin and installed on the jobsite in relatively proximity to the machines). Two different jobsite conditions have also been considered: 1) jobsites with single machines and 2) jobsites with multiple machines operating simultaneously. In terms of software settings, two different SVM classifiers (RBF and Linear Kernels) and two common frequency feature extraction techniques (STFT and CWT) were selected and evaluated. Experimental data was gathered and used to optimize hardware settings, tune algorithmic parameters, and evaluate different approaches. Results depict an accuracy over 85% for the proposed audio-based recognition system.

Key Words: Microphone arrays; audio signal processing; construction equipment; activity recognition; Fourier Transform.

INTRODUCTION

Recognizing the activities of construction heavy equipment is a major step toward productivity analysis of a jobsite. Aside from productivity monitoring, recognizing the activities of construction machinery provides useful information for various other purposes including scheduling and cost estimation (Rashidi, Nejad and Maghiar 2014), job site layout analysis and decision making (Pradhananga and Teizer 2013), and monitoring and optimization of fuel consumption (Akhavian and Behzadan 2013a). Some manufacturers have started to integrate their proprietary monitoring devices to new equipment, but construction contractors are unlikely to have a fleet composed entirely by new and single-manufacturer equipment for this to represent an immediate solution. The common alternative approaches for heavy equipment activity recognition require active and/or passive external sensors (Ahn, Lee and Peña-Mora 2015, Akhavian and Behzadan 2013b, Akhavian and Behzadan 2015, Brilakis, Fathi and Rashidi 2011, Golparvar-Fard,

Heydarian and Niebles 2013, Gong, Caldas and Gordon 2011, Heydarian, Golparvar-Fard and Niebles 2012, Kim and Caldas 2013, Rashidi, Fathi and Brilakis 2011, Rezazadeh Azar and McCabe 2012, Zhu, et al. 2016). Active sensors include GPS and micro-electro-mechanical systems (MEMS) devices such as accelerometers and gyroscopes, while passive sensors include image/video processing using computer vision algorithms (National Instruments 2016).

During the past few years, another category of input data, audio, has been investigated by researchers for recognizing various activities of construction heavy equipment and machines. The idea is that construction heavy equipment generates distinct sound patterns while performing different activities that can provide useful information when properly recorded and processed. gap (Cheng, et al 2017; Cho, et al 2017; Yang and Cao 2015, Cao et al, 2016). The idea of analyzing sounds generated by various engineering systems and devices has appeared in several research and application areas such as speech recognition, audio-based navigation of robots, the use of sonar in the exploration and mapping, ultrasonic signal processing for condition based maintenance (CBM) in manufacturing workplaces (Bengtsson, et al. 2004, Greenemeier 2008), and audio signal processing for feature recognition in medical devices. Each of these applications require their own hardware and software settings (types of microphone and other acoustic recording devices, layout of microphones, number and distance between the sound capturing devices, etc.)

Despite the widespread use of audio signal processing techniques for analysis and modeling of various engineering techniques, applications of these methods in construction management area is still in early stages of development. This paper summarizes recent studies conducted by the authors identifying necessary hardware devices and computing techniques, mostly suitable for capturing and processing audio files in large scale, noisy, and cluttered construction jobsites.

LITERATURE REVIEW: ACOUSTICAL MODELING OF ENGINEERING SYSTEMS

Bengtsson, et al. (2004) define audio-based feature identification as a three-module process (Fig. 1). First, the sound is recorded into a computer as an input to the processing module. The processing module consists of two steps: pre-processing to remove unwanted noise and to extract a period of interest and feature extraction to identify characteristic features of the audio sample and create a feature vector. Once a feature vector is created, the condition monitoring and diagnosis module consists in comparing the audio sample to an existing library and providing a condition diagnosis. If the software is in the learning mode, newly-identified vectors are stored into the case library. Although this model describes ultrasound-based CBM systems, it is characteristic to most audio-based machine learning practices and can be applied to other fields.

Some typical applications of ultrasound-based CBM include: bearing inspection; testing gears/gearboxes; pumps; motors; steam trap inspection; valve testing; detection/trending of cavitation; compressor valve analysis; leak detection in pressure and vacuum systems such as boilers, heat exchangers, condensers, chillers, tanks, pipes, hatches, hydraulic systems, compressed air audits, specialty gas systems and underground leaks; and testing for arcing and corona in electrical apparatus (Naik 2009).

In the medical setting, similar technologies have been developed to monitor patients with chronic disease such as asthma and chronic obstructive pulmonary disease (COPD). Specifically, to evaluate the adherence to inhaler medication, which involves doses being taken in a consistent schedule and with a proper method. The inhaler compliance assessment (INCA) device was developed as an approach to evaluate inhaler adherence. This device is attached to a widely-used variety of inhalers, as shown in Fig. 2 (right). When the patient opens the mouthpiece to take a

dose, the INCA device takes an audio recording with a timestamp. An example of a typical audio sample after applying the fast Fourier transform (FFT) algorithm to extract its frequency features is shown in Fig. 2 (left). One month of typical inhaler use yields 60 audio files corresponding to 60 doses of medication. Analyzing this data set takes an experienced pulmonary clinician an average of 30 minutes. This type of labor intensive analysis is not feasible with a large group of patients. Thus, a computer algorithm has been designed to analyze audio data sets and provide an adherence score based on dose schedule and the pattern of blister, exhalation, and inhalation events (Holmes, et al. 2014). This sort of study provides useful information for clinicians to understand why a specific inhaler is not effective and propose methods to enhance the patient's adherence (Sulaiman, et al. 2017).

Although the potential for applications is boundless, there has been little prior work studying the implementation of audio signal processing techniques specifically in the area of construction engineering and management. The motivation behind an audio based activity recognition framework for construction operations lies in the inherent limitations of the existing techniques for detailed assessment of construction activities. These limitations have hindered the successful adoption of these techniques and hence the industry still relies heavily on the experience of project managers on the jobsite to perform such assessments. In comparison with other active and passive activity recognition methods, working with audio signals provides the following advantages:

- Most active sensors (GPS, accelerometers, etc.) need to be directly mounted on the equipment. In addition, for each machine, at least one sensor is required. As explained in this work, this is not a limitation for an audio-based activity recognition method. Microphones

can be installed in various locations at the jobsite and one microphone is usually able to cover the activities of many machines.

- Computer vision methods are very sensitive to environmental factors such as lighting conditions and occlusions. In addition, the limited field of view of cameras is another major drawback for computer vision methods. Microphones and audio signals are more resilient against the above-mentioned limitations.
- The sound produced by each piece of equipment is independent of the operator. As the result, an operator can perform an activity in a variety of ways. An action recognition system based on location sensors and/or computer vision techniques would need to take these different scenarios into account, while an audio analysis always yields the same result.
- Compared to video data, audio files have lower data rates and are therefore computationally more efficient.

Choosing the optimal hardware setting (type and location of microphones) and software (selecting proper algorithms and tuning algorithmic parameters) is a crucial stage for acoustical modeling of real-world construction jobsites.

HARDWARE SETTING: DIFFERENT TYPES OF MICROPHONES

The primary devices used for audio recording are microphones. A microphone generally contains a diaphragm, either surface or moving, designed to capture electroacoustic waves and generate electronic signals. Typical microphone classifications involve distinguishing either by their pickup pattern or by the type of transducer they employ.

Per Ballou (2015), pickup pattern refers to how the microphone discriminates among different directions of the incoming sound. The most common microphone types are

omnidirectional, bidirectional, and unidirectional or cardioid microphones. In an omnidirectional microphone, the pickup pattern is equal in all directions. Omnidirectional microphones are particularly useful when it is necessary to capture all audio sources in an environment. In a bidirectional microphone, the pickup pattern is equal in two opposite directions and negligible at 90° from these two directions. Bidirectional microphones are particularly useful when it is necessary to capture the conversations of two speakers positioned face to face. In a unidirectional microphone, the pickup pattern has a cardioid shape facing in just one direction. Of all three, unidirectional microphones are the most widely used because they allow the user to focus on a specific source of interest. Pickup pattern is a key factor for acoustical modeling of construction jobsites since, in a multifaceted jobsite, several pieces of construction equipment might work simultaneously, generating sound from multiple locations and directions. More details of the aforementioned microphones plus some variations of these are provided in Fig. 3 (Ballou 2015).

A transducer is a device that converts a physical stimulus to an electrical signal output. Transducers can be carbon, crystal, and ceramic microphones, condenser microphones, dynamic microphones, and electret condenser microphones. Microphone selection by type of transducer is another key consideration in a construction jobsite. Transducers must be robust enough to withstand inclement weather and other contingencies while maintaining stable audio signal recording characteristics.

For this research three microphones of varying types have been selected (Fig. 4): 1) a Zoom H1 digital handy recorder, an off-the-shelf microphone; 2) a Korg CM-200 clip-on contact microphone; and 3) an xCore-200 multichannel microphone array. Condenser microphones and a variant of these called MEMS microphones are the type of microphone built into the Zoom H1

handy recorder and the XMOS microphone array, respectively. Per Yamaha (2016), condenser microphones have good sensitivity to all frequencies, but are highly susceptible to structural vibration and humidity. Crystal and ceramic microphones are the type of microphone built into the Korg contact microphones (Korg 2016). Their name comes from the fact that these devices use piezoelectric crystals such as Rochelle salt, tourmaline, barium titanate, and quartz to convert pressure differences from the environment into a voltage output. Due to this pressure response characteristic, crystal microphones can be built to pick up vibrations from contact with objects, as opposed to common microphones that pick-up air-carried sound waves.

Contact microphones are built with a unidirectional, piezoelectric transducer, which is designed to be less susceptible to air-carried sound waves and more susceptible to surface-borne sound waves. The Korg CM-200 microphone, selected for data collection in this research, is commonly used in applications that involve capturing sound from a particular musical instrument (e.g., brass instrument, guitar, violin, and ukulele) when recording or practicing with an entire music band (Korg 2016). Contact microphones have the potential to be attached to heavy machinery while not being overly susceptible to the vibrations of the machine itself, unlike condenser microphones.

A microphone array is composed of at least two microphones working in tandem arranged in a linear, rectangular, or circular pattern. These microphones are usually omnidirectional; however, some microphone arrays are built using directional microphones or a combination of omnidirectional and directional microphones. Brandstein and Ward (2001) compiled over 20 years of research in array-based microphone technology, which resulted in a thorough reference for immediate applicability of array technology in current systems and in improvement of existing

devices. The applications for array-based microphones recorded in this compilation are based on beamforming techniques and include speech enhancement, speech recognition, source localization, noise reduction, echo cancellation, and separation of acoustic signals. Source localization and separation of acoustic signals have the potential to be very useful tools for locating heavy equipment and separating audio signals from other machinery working simultaneously in cluttered jobsites.

The XMOS xCORE-200 microphone array board, selected for data collection in this research, is a hardware and reference software platform equipped with seven omnidirectional MEMS microphones with pulse density modulation (PDM) output (one microphone is placed at the center at the board and the remaining six microphones are distributed equidistantly around the board edge, as shown in Fig. 4), a digital-analog converter (DAC), a processor with sixteen 32-bit logical cores, on-board low-jitter clock sources for multiple clocking options, four configurable buttons, 13 LED indicators, a USB 2.0 port, a RJ45 Ethernet port, a 3.5 mm audio jack, and other components. This particular microphone array board and its reference developer firmware are targeted, but are not strictly limited to, Voice User Interface (VUI) applications (XMOS Ltd. 2016).

RESEARCH METHODOLOGY: AUDIO-BASED ACTIVITY RECOGNITION OF CONSTRUCTION HEAVY EQUIPMENT

To determine the software and hardware requirements for acoustical modeling of construction jobsites, the authors have developed an audio-based model for activity recognition of construction heavy equipment (Fig. 5). The five core components will be briefly described below.

Hardware settings for data collection will be discussed thoroughly in the next section. More detailed information about the implemented system can be found at (Cheng, et al. 2016).

Denoising

Audio samples of heavy equipment mixed are almost certainly contaminated with noise from other sound sources found in a job site. To reduce the effect of this noise on later processing tasks, it is necessary to filter out as much noise as possible prior to subsequent processing. However, noise filtering can also degrade the desired (equipment) signal and so the filtering must be balanced to reduce unwanted noise effectively while minimizing the distortion of the sound patterns of interest. Additionally, noise filtering must remain effective under the assumption that noise sources at a job site are not necessarily constant since workers and equipment perform short tasks in an intermittent manner. Thus, a denoising algorithm for non-stationary environments developed by Rangachari and Loizou (2006) was selected for implementation using MATLAB. This algorithm uses a noise estimation method for non-stationary environments that rapidly adapts depending on relative level of noise versus signal of interest.

Sound Source Detection (Steered Response Power and Beamforming)

The case of dealing with multiple machines operating simultaneously at a jobsite is more challenging due to overlap between the sounds patterns generated by various machines. As a result, it is necessary to separate generated sound patterns by each individual machine first. For multiple machine recordings, sounds from different machines come from different directions. A robust source detection and activity recognition algorithm must respond to sound from a specific direction and block most of the noise from outside the direction of interest. To achieve this goal, the first

step is to estimate the arrival direction for the sounds of interest. Direction of arrival for different machines is estimated using SRP (Steered Response Power). The idea of SRP is that the location of the signal source radiates more energy than all other locations (Dmochowski, Benesty and Affes 2007). Thus, we can identify directions of arrival by scanning over all feasible angles and choosing the one with the largest value power. In practice, we scan through a hemisphere and find the direction with largest power. After finding the range, elevation and azimuth value from SRP, beamforming is used to isolate the signal from the desired direction. Beamforming consists of a spatial filter that uses the inputs from multiple microphones to isolate (as much as possible) signals arriving from a particular direction (Gannot, Burshtein and Weinstein 2001, Johnson and Dudgeon 1993). The illustration of ideal direction detection result is shown in Fig.6. The top figure shows the positions for microphones and sound sources, and the bottom figure shows the directions detected by the SRP technique. In our pipeline, we choose delay-and-sum beamformer since it is the simplest and still highly effective.

Short-Time Fourier Transform (STFT)

Once enhanced and isolated, each audio signal is then converted into a time-frequency domain representation using the STFT. This technique consists of dividing a temporal signal into short segments and computing the Fourier transform for each segment; thereby extracting sinusoidal frequency, magnitude, and phase content of each segment and representing these features as they change over time. MATLAB was used for STFT implementation using a Hanning widow size of 512, a 1024-point discrete Fourier transform, and a 50% overlap. For the subsequent processing, only the magnitude content is retained.

Support Vector Machines (SVMs)

To identify activities being performed by construction equipment within an audio recording, an SVM classifier was trained to recognize each piece of equipment while performing major activities (class 1) and minor activities (class 2). The LIBSVM MATLAB package was used for this task (Chang and Lin 2011). Four audio segments of the construction equipment performing a major activity are used to train class 1, and four audio segments of the construction equipment performing a minor activity were used to train class 2. Each of these segments was selected to be 2 to 6 seconds long and included only the STFT magnitude. To guarantee correct SVM parameter selection, ten-fold cross validation was used. It is well known that the performance of the SVM algorithm highly depends on the selected kernel function (Rashidi, Sigari, et al. 2016). For this research, the linear and radial basis kernel functions have been selected, and experiments have been performed with both.

Window Filtering

Once an SVM classifier is generated for a specific construction equipment, it can be used for classification of the rest of the audio file. Nonetheless, direct implementation would potentially yield an output with predicted activities changing erratically from one time-frequency bin to the next. Therefore, a window filtering algorithm was implemented to smooth out the classified output. The window filtering parameters are a small window size, a large window size, and a threshold. Initially, if the SVM labels indicate that the percentage for a certain activity is greater than the threshold throughout the small window, the whole small window is labeled as that activity. Then, this is repeated using the small window labels for the large window size. The size of the window

will vary in different cases, but in general the small window can be set as a quarter second and the large window can be a second or two seconds.

EXPERIMENTAL SETUP: Audio-Based Activity Recognition of Single Machines

A selection of various construction heavy equipment has been chosen to help assess how well the audio-based activity analysis system performs under different hardware and software settings. Using four recording devices, audio data from heavy machinery performing routine activities was collected. The selected microphones were two off-the-shelf Zoom H1 handy recorders (regular microphone), one on a tripod located on-site and one on-board the machine; one XMOS xCORE-200 USB microphone array connected to a laptop; and one KORG CM-200 contact microphone connected to a flat surface in the cabin of the equipment. The XMOS microphone array was interfaced to a Windows PC using the manufacturer's USB Audio Class 2.0 Evaluation Driver for Windows and multichannel audio was recorded through Audacity, an open source program for recording and processing audio. The setup process for data collection is depicted in Fig. 7.

EXPERIMENTAL SETUP: Audio-Based Activity Recognition of Multiple Machines Working Simultaneously on a Jobsite

The experimental setup for the single machine case is almost the same as the multiple machine case. The only difference is that we only use the XMOS xCORE-200 USB microphone array and regular on-site microphone in the multiple machine cases. Also, we record each machine separately to be used as our training library in the multiple machine cases. Recordings from the microphone array can provide us with phase information, which can help us track the location of

sound sources as described above. With the spatial information, the proposed activity recognition framework can work better in the multiple machine cases.

In addition to audio data recording, a video sample was taken using a digital video camera to serve as a reference for manual labeling of major and minor activities. Examples of major activities include digging, loading, dumping, and crushing rock, while examples of minor activities include swinging, maneuvering, and extending arm. Once the devices were in place and recording, a loud sound was produced with an air horn. This signal was to be used as a synchronization point for the data. Manual labels served as ground truth data to assess the activity recognition accuracy.

The results for one sample machine are depicted in Fig. 8. The machine, an Ingersoll Rand SD25 compactor, was recorded using a Zoom H1 digital handy recorder placed relatively close to the machine on the jobsite. The presented results are based on using an SVM with a linear kernel. The top part of the figure presents the normalized frequency for the machine's audio signal, while the middle and bottom plots show the actual (black) versus predicted (blue) activity labels and over the audio recording time period. As indicated in these figures, there is an excellent correlation between the actual and predicted results generated for the Ingersoll Rand compactor.

EXPERIMENTAL SETUP: Comparison between Different Frequency Feature Extraction Techniques

Selecting optimum frequency feature extraction method is an important decision for optimizing the performance of the proposed audio-based activity recognition package. In this study, two common feature extraction techniques, the short-time Fourier transform (STFT) and the continuous wavelet transform (CWT) have been selected and compared. STFT is based on dividing a long-time signal into short segments and computing the Fourier transform for each

segment. CWT is a derivation of the Fourier transform that was primary designed to locate frequency features in time or space. The Fourier transform is a powerful tool for frequency analysis; however, it does not characterize rapid frequency changes efficiently because it represents data as a sum of sine waves, which extend to infinity. A wavelet is a rapidly-decaying, wave-like oscillation that has zero mean. More detailed information regarding each feature extraction method could be found at relevant references such as Mathwoks, Inc. report, 2017.

To conduct the comparison study between STFT and CWT, both techniques were implemented using recordings taken at local construction jobsites for various types of equipment. Two configurations for CWT (8 octaves/ 32 scales per octave; and 10 octaves and 24 scales per octave) and one configuration for STFT (1024 frequency points, 512-sample window, and 256 overlapped samples) have been employed in this research.

RESULTS AND DISCUSSION

Tables 1 – 5 and Figures 9-11 summarize comparison results for several construction equipment under different hardware and software configurations. Through careful analysis of these results, we arrive at the following conclusions:

- For on-board microphone placement, contact microphones generate slightly more accurate results and, thus, are the better choice (Table 1).
- A comparison for the current case study results shows that regular microphones placed on the construction site yield better accuracies than contact microphones attached to flat surfaces in the cabin. One clear reason for this phenomenon is that contact microphones on-board are affected by engine noise and the vibrations; nonetheless, further investigations using several

other data sets and under different jobsite conditions might be required to fully substantiate this conclusion (Tables 2 and 3).

- As shown in Fig. 9, results show that the radial basis function kernel outperforms the linear kernel in most cases, particularly when it comes to recognizing the major activities.
- No significant differences can be found between regular microphones and microphone arrays when it comes to recognizing the activities of single machines through microphones placed on-site. This is due to there being only one source of audio, and therefore only one direction of incoming audio. It is suggested that a noticeable difference between a regular microphone and a microphone array will be seen in the recording of multiple machines, and the multiple directions of generated audio from said machines. Microphone arrays will become useful to detect multiple audio directions from multiple audio sources.
- In Table 4, we can find in the multiple machine case, microphone array can provide us with better recognition results compared to regular on-site microphone. The reason is that with direction detection step, we can eliminate most of the noise outside the direction of interest. When using the microphone array, we improve the performance of activity recognition by using recordings that only contain one machine sound as our training library. After the direction detection step, we used the trained library to perform the activity recognition in the mixed recording. However, when using a single on-site microphone, we can only do the activity recognition on the mixed recording since we are lacking spatial information.
- In general, it can be observed that CWT scalograms yield a better time-frequency magnitude representation than STFT spectrograms. Additionally, it has been illustrated that using a smaller number of octaves with higher resolution within the octaves is preferable when lower-

frequency features are not of interest. Using a CWT with 8 octaves outperformed using a CWT 10 with octaves on processing time and labeling accuracy.

SUMMARY AND CONCLUSION

This work provides an innovative audio-based solution for activity analysis of construction heavy equipment and acoustical modeling of construction jobsites. An audio-based activity recognition model has been developed and tested under various hardware and software settings for the case involving single machines. The methods of the experiment apply the three types of microphones discussed earlier (off-the-shelf, contact, and microphone arrays) and two placement settings (on-board vs. on-site). The comparison results have been presented in the previous sections of this paper. Additionally, the results were also processed with two major SVM kernel types (linear and RBF) to determine which kernel yielded the best results. The RBF kernel performed better in most cases. As an extension of the current research, plans for future research include expanding into the following items:

- Implementing and testing the audio-based model with more diverse data sets and conditions
- Determining what hardware and software settings would be required to implement this system on a jobsite with multiple machines working simultaneously
- Developing more robust algorithms that would recognize more detailed tasks performed by heavy equipment and machinery. That is, separating major and minor activities into sub-activities.

ACKNOWLEDGEMENTS

The presented research has been funded by the U.S. National Science Foundation (NSF) under Grants CMMI-1606034 and CMMI-1537261. The authors gratefully acknowledge NSF's support. Any opinions, findings, conclusions, and recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the NSF.

References

- Ahn, C., S. Lee, and F. Peña-Mora. 2015. "Application of low-cost accelerometers for measuring the operational efficiency of a construction equipment fleet." *Journal of Computing in Civil Engineering* 29 (2): 04014042.
- Akhavian, R., and A. H. Behzadan. 2015. "Construction equipment activity recognition for simulation input modeling using mobile sensors and machine learning classifiers." *Advanced Engineering Informatics* 29 (4).
- Akhavian, R., and A. H. Behzadan. 2013b. "Knowledge-based simulation modeling of construction fleet operations using multimodal-process data mining." *Journal of Construction Engineering and Management* 139 (11): 04013021.
- . 2013a. "Simulation-based evaluation of fuel consumption in heavy construction projects by monitoring equipment idle times." *Simulation Conference (WSC)*.
- Ballou, Glen. 2015. *Handbook for Sound Engineers*. 5th. New York: Focal Press.
- Bengtsson, M., Olsson, Funk, P. E., and M. Jackson. 2004. "Technical design of condition based maintenance system—A case study using sound analysis and case-based reasoning." *International Conference of Maintenance and Reliability (MARCON)*. Knoxville, TN.
- Brandstein, Michael, and Darren Ward,. 2001. *Microphone Arrays: Signal Processing Techniques and Applications*. Springer.

- Brilakis, I., H. Fathi, and A. Rashidi. 2011. "Progressive 3D reconstruction of infrastructure with videogrammetry." *Automation in Construction* 20 (7): 884-895.
- Cao, J., W. Wang, J. Wang, and R. Wang. 2016. "Excavation equipment recognition based on novel acoustic statistical features." *IEEE Transactions on Cybernetics (IEEE)* 47 (12): 4392 - 4404.
- Chang, C.C., and C.J. Lin. 2011. "LIBSVM: a library for support vector machines." 2 (3): 1-27.
- Cheng, C. F., A. Rashidi, and D. V. Anderson M. A. Davenport. 2017. "Activity analysis of construction equipment using audio signals and support vector machines." *Automation in Construction* 81.
- Cheng, C., F., A. Rashidi, M. Davenport, and D. Anderson. 2016. "Audio Signal Processing for Activity Recognition of Construction Heavy Equipment." *33rd International Symposium on Automation and Robotics in Construction (ISARC 2016)*. Auburn, AL.
- Cho, Chunhe, Yong-Cheol Lee, and Tianyi Zhang. 2017. "Sound Recognition Techniques for Multi-Layered Construction Activities and Events." *ASCE International Workshop on Computing in Civil Engineering*. Seattle: American Society of Civil Engineers. 326-334.
- Dmochowski, Jacek P., Jacob Benesty, and Sofiene Affes. 2007. "A generalized steered response power method for computationally viable source localization." 15 (8): 2510-2526.
- Gannot, Sharon, David Burshtein, and Ehud Weinstein. 2001. "Signal enhancement using beamforming and nonstationarity with applications to speech." *IEEE Transactions on Signal Processing* 49 (8): 1614-1626.

- Golparvar-Fard, M., A. Heydarian, and J.C. Niebles. 2013. "Vision-based action recognition of earthmoving equipment using spatio-temporal features and support vector machine classifiers." *Advanced Engineering Informatics* 27 (4): 652-663.
- Gong, J., C.H. Caldas, and C. Gordon. 2011. "Learning and classifying actions of construction workers and equipment using Bag-of-Video-Feature-Words and Bayesian network models." *Advanced Engineering Informatics* 25 (4): 771-782.
- Greenemeier, L. 2008. *A positioning system that goes where GPS can't*. Scientific American. January 23. <http://bit.ly/2hsVPNJ>.
- Heydarian, A., M. Golparvar-Fard, and J. C. Niebles. 2012. "Automated Visual Recognition of Construction Equipment Actions Using Spatio-Temporal Features and Multiple Binary Support Vector Machines." *Construction Research Congress*. ASCE. 889-898. doi:10.1061/9780784412329.090.
- Holmes, Martin S., Shona D'Racy, Richard W. Costello, and Richard B. Rielly. 2014. "Acoustic Analysis of Inhaler Sounds from Community-Dwelling Asthmatic Patients for Automatic Assessment of Adherence." *IEEE Journal of Translational Engineering in Health and Medicine* (IEEE) 2. doi:10.1109/JTEHM.2014.2310480.
- Johnson, Don H., and Dan E. Dudgeon. 1993. *Array Signal Processing: Concepts and Techniques*. Englewood Cliffs, New Jersey: Prentice Hall.
- Kim, J Y, and C H Caldas. 2013. "Vision-Based Action Recognition in the Internal Construction Site Using Interactions between Worker Actions and Construction Objects." *30th International Association for Automation and Robotics in Construction*. International Association for Automation and Robotics in Construction. 661-668.

- Korg. 2016. *CM-200 Contact Microphone*. Accessed October 8, 2016. <http://bit.ly/2qAhUOf>.
- MathWorks, Inc. 2017-a. *Time-Frequency Analysis with the Continuous Wavelet Transform*. Accessed October 2017. <http://bit.ly/2yyMM7j>.
- Naik, R. P. 2009. *Ultrasonic: A new method for condition monitoring*. Raipur, National Thermal Power Corporation. 10 12. Accessed 03 26, 2016. <http://www.reliableplant.com/Read/20554/ultrasonic-condition-monitoring>.
- National Instruments. 2016. *What Determines if a Transducer Is Active or Passive?* March 12. Accessed September 30, 2016. <http://bit.ly/1CSIyS3>.
- Pradhananga, Nipesh, and Jochen Teizer. 2013. "Automatic spatio-temporal analysis of construction site equipment operations using GPS data."
- Rangachari, S., and P Loizou. 2006. "A noise estimation algorithm for highly nonstationary environments." *Speech Communication* 28: 220-231. doi:10.1061/(ASCE)CO.1943-7862.0000652.
- Rashidi, A., H. Fathi, and I. Brilakis. 2011. "Innovative stereo vision-based approach to generate dense depth map of transportation infrastructure." *Transportation Research Record: Journal of the Transportation Research Board* 2215: 93-99.
- Rashidi, A., H.R. Nejad, and M. Maghiar. 2014. "Productivity estimation of bulldozers using generalized linear mixed models." *KSCE Journal of Civil Engineering* 18 (6): 1580-1589.
- Rashidi, A., M.H. Sigari, M. Maghiar, and D. Citrin. 2016. "An analogy between various machine learning techniques for building materials recognition in construction site images." *KSCE Journal of Civil Engineering* 20 (4): 1178-1188.

- Rezazadeh Azar, E., and B. McCabe. 2012. "Part based model and spatial–temporal reasoning to recognize hydraulic excavators in construction images and videos." *Automation in Construction* 24: 194-202.
- Sulaiman, I., B. Cushen, G. Greene, J. Seheult, D. Seow, F. Rawat, E. MacHale, et al. 2017. "Objective Assessment of Adherence to Inhalers by Patients with Chronic Obstructive Pulmonary Disease." *American Journal of Respiratory and Critical Care Medicine* 195 (10). doi:10.1164/rccm.201604-0733OC.
- XMOS Ltd. 2016. "xCORE Microphone Array Hardware Manual." March 2.
- Yamaha. 2016. *Which types of Microphone Are Used with PA systems?* April 2. <http://bit.ly/2cMoi0Y>.
- Yang, S., and J. Cao. 2015. "Acoustics recognition of construction equipments based on LPCC features and SVM." *34th Chinese in Control Conference*. IEEE. 3987-3991.
- Zhu, Zhenhua, Man-Woo Park, Christian Koch, Mohamad Soltani, Amin Hammad, and Kheshayar Davari. 2016. "Predicting movements of onsite workers and mobile equipment for enhancing construction site safety." *Automation in Construction* 95-101. doi: 10.1016/j.autcon.2016.04.009.
- Zoom Corporation. 2016. Accessed May 13, 2017. <https://www.zoom.co.jp/products/handy-recorder/h1-handy-recorder>.

List of Tables

TABLE 1. Comparison between contact and regular microphones (on-board/SVM using the RBF kernel)

TABLE 2. Comparison between contact (on-board) and regular (on-site) microphones.

TABLE 3. Comparison between on-site and on-board regular microphones.

TABLE 4. Comparison between array and on-site regular microphones on multiple machine recordings

TABLE 5. Comparison between array and on-site regular microphones on multiple machine recordings

TABLE 1. Comparison between contact and regular microphones (on-board/ SVM using the RBF kernel).

Machine	Contact Microphone		Regular microphone	
	Major activity	Minor activity	Major activity	Minor activity
JD50D Compact Backhoe	72.08%	74.02%	71.14%	82.78%
Ingersoll Rand Compactor	80.34%	7.83%	80.07%	1.98%
CAT 320E Excavator	80.78%	38.26%	71.20%	29.71%
Komatsu PC200 Excavator	71.05%	48.56%	70.16%	55.59%
JD 700J Dozer	76.16%	64.46%	78.79%	60.15%
Hitachi 50U Excavator	49.58%	57.96%	53.15%	54.17%
Concrete Mixer 2	68.79%	65.68%	77.16%	70.23%

TABLE 2. Comparison between contact (on-board) and regular (on-site) microphones.

Machine	Contact Microphone (on-board)		Regular microphone (on-site)	
	Major activity	Minor activity	Major activity	Minor activity
CAT 320D Backhoe Excavator	80.78%	38.26%	81.09%	71.48%
JD 333E Compact Loader	73.42%	95.68%	86.54%	90.11%
JD50D Compact Backhoe	72.08%	74.02%	86.65%	57.74%
Ingersoll Rand Compactor	80.34%	7.83%	81.58%	32.03%
CAT 320E Excavator	80.78%	38.26%	81.09%	71.48%

Komatsu PC200 Excavator	71.05%	48.56%	82.17%	67.89%
JD 700J Dozer	76.16%	64.46%	84.69%	72.45%
Concrete Mixer 2	68.79%	65.68%	83.23%	61.59%

TABLE 3. Comparison between on-site and on-board regular microphones.

Machine	On-site		On-board	
	Major activity	Minor activity	Major activity	Minor activity
JD50D Compact Backhoe	86.65%	57.74%	71.14%	82.78%
Ingersoll Rand Compactor	81.58%	32.03%	80.07%	1.98%
JD 333E Compact Loader	86.54%	90.11%	80.67%	81.47%
CAT 320E Excavator	81.09%	71.48%	71.20%	29.71%
Komatsu PC200 Excavator	82.17%	67.89%	70.16%	55.59%
JD 700J Dozer	84.69%	72.45%	78.79%	60.15%
Hitachi 50U Excavator	79.89%	55.97%	53.15%	54.17%
Concrete Mixer 2	83.23%	61.59%	77.16%	70.23%

TABLE 4. Comparison between array and on-site regular microphones on multiple machine recordings.

Multiple machine cases		Microphone Array		On-site regular microphone	
		Major activity	Minor activity	Major activity	Minor activity
Bobcat Loader and Bobcat Excavator	Bobcat Loader	72.37%	52.74%	65.34%	62.78%
	Bobcat Excavator	74.29%	60.17%	68.16%	51.17%
Bomag Compactor and JD 544K Loader	Bomag Compactor	81.12%	49.07%	70.22%	57.88%
	JD 544K Loader	76.14%	55.47%	59.74%	41.81%
	CAT Bulldozer	80.42%	81.33%	71.54%	65.14%

CAT Bulldozer and CAT CS44 Compactor	CAT CS44 Compactor	79.08%	77.29%	67.44%	50.08%
CAT C5K Bulldozer and CAT 305E Excavator	CAT C5K Bulldozer	78.57%	61.90%	57.82%	39.97%
	CAT 305E Excavator	77.69%	64.18%	62.26%	47.64%

TABLE 5. Comparison between CWT 10/24 vs. STFT true positive classification accuracy.

	CWT 10/24		STFT	
	Act 1	Act 2	Act 1	Act 2
JD 670G	91.76%	52.46%	79.31%	79.28%
JCB 3CX	82.06%	60.73%	91.44%	54.62%
Komatsu PC200	69.45%	68.34%	62.18%	72.75%

List of Figures

Fig. 1. Schematic for computer-based feature identification (Bengtsson, et al. 2004)

Fig. 2. Spectrogram of audio signal caught with INCA device showing blister, exhalation, and inhalation events (left); INCA device attached to inhaler (right) – (Holmes, et al. 2014)

Fig. 3. Performance characteristics of several microphones (Ballou 2015)

Fig. 4. Different types of microphones used in the project: Zoom H1 digital handy recorder (left) (Zoom Corporation 2016); Korg CM-200 clip-on contact microphone attached to construction equipment (center); xCORE-200 microphone array evaluation board – top view – (right) (XMOS Ltd. 2016).

Fig. 5. The audio-based model for activity analysis of construction heavy equipment.

Fig. 6. Direction Detection Steps; The positions for microphone array and sound sources (left); Two directions detected by steered response power method (right)

Fig. 7. Setup process for audio collections using microphone array and off-the-shelf microphone on-site (left), and contact microphone and off-the shelf microphone on-board (middle); Simultaneous video recording for generating the ground truth data (right)

Fig. 8. Results for Ingersoll Rand SD25 Compactor.

Fig. 9. Summary of the results for implementing RBF and Linear Kernels for activity recognitions of machines (major activities)

Fig. 10. Sample of comparison results between CWT (10/24) vs. STFT spectrograms- Machine: JCB 3CX

Fig. 11. Labeling comparison between CWT (10/24) vs. STFT spectrograms- Machine: JCB 3CX

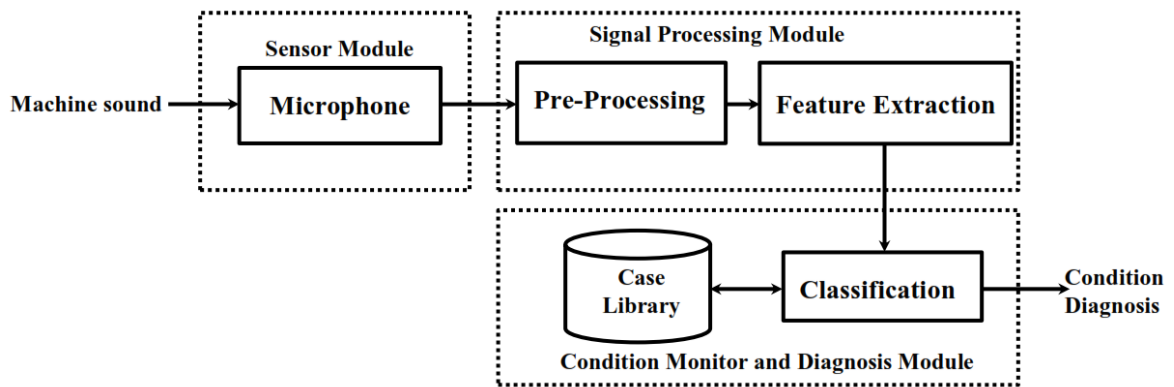


Fig. 1. Schematic for computer-based feature identification.

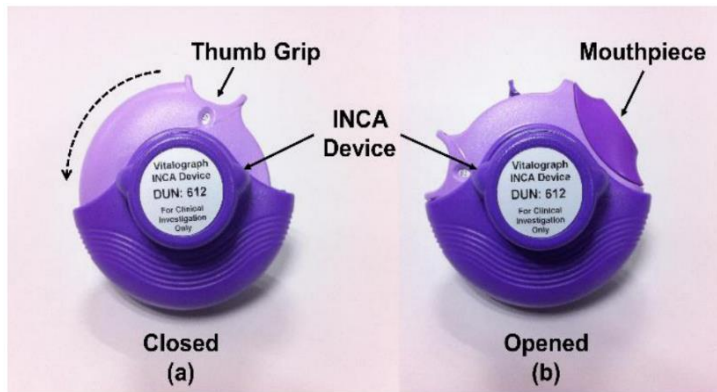
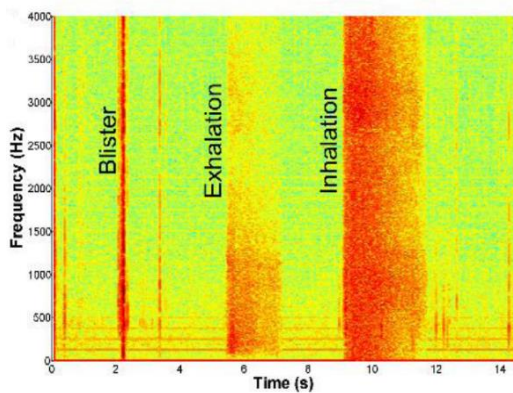


Fig. 2. Spectrogram of audio signal caught with INCA device showing blister, exhalation, and inhalation events (left); INCA device attached to inhaler (right) – (Holmes, et al. 2014).


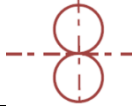



Microphone	Omnidirectional	Bidirectional	Directional	Supercardioid	Hypercardioid
Directional response characteristics					
Voltage output	$E = E_o$	$E = E_o \cos \theta$	$E = \frac{E_o}{2}(1 + \cos \theta)$	$E = \frac{E_o}{2}[(\sqrt{3} - 1) + (3\sqrt{3}) \cos \theta]$	$E = \frac{E_o}{4}(1 + 3 \cos \theta)$
Random energy efficiency (%)	100	33	33	27	25
Front response	1	1	∞	3.8	2
Back response					
Front random response	0.5	0.5	0.67	0.93	0.87
Total random response					
Front random response	1	1	7	14	7
Back random response					
Equivalent distance	1	1.7	1.7	1.9	2
Pickup angle (2θ) for 3 dB attenuation	-	90°	130°	116°	100°
Pickup angle (2θ) for 6 dB attenuation	-	120°	180°	156°	140°

Fig. 3. Performance characteristics of several microphones.

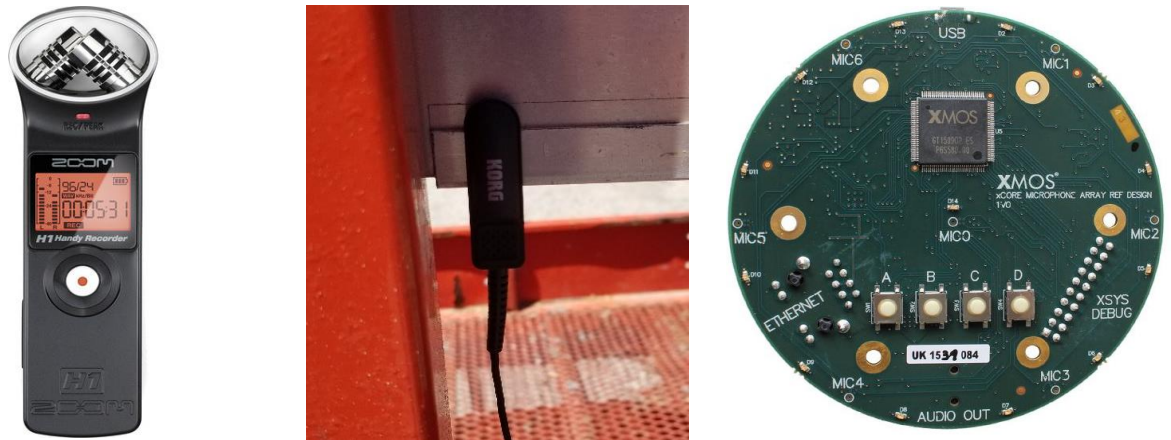


Fig. 4. Different types of microphones used in the project: Zoom H1 digital handy recorder (left) (Zoom Corporation 2016); Korg CM-200 clip-on contact microphone attached to construction equipment (center); xCORE-200 microphone array evaluation board – top view – (right) (XMOS Ltd. 2016).

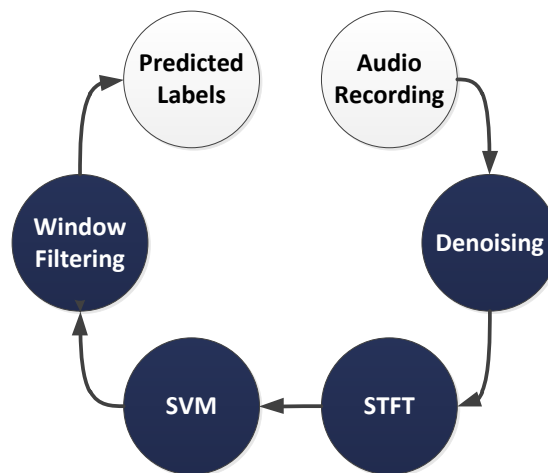


Fig. 5. The audio-based model for activity analysis of construction heavy equipment.

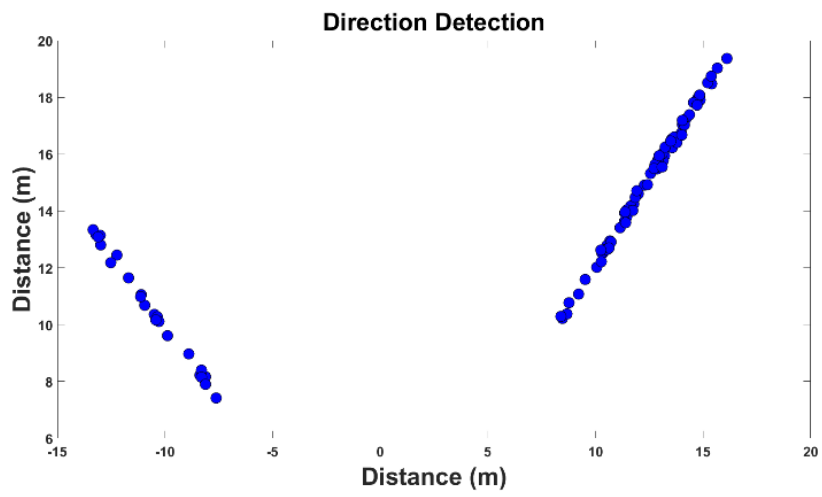
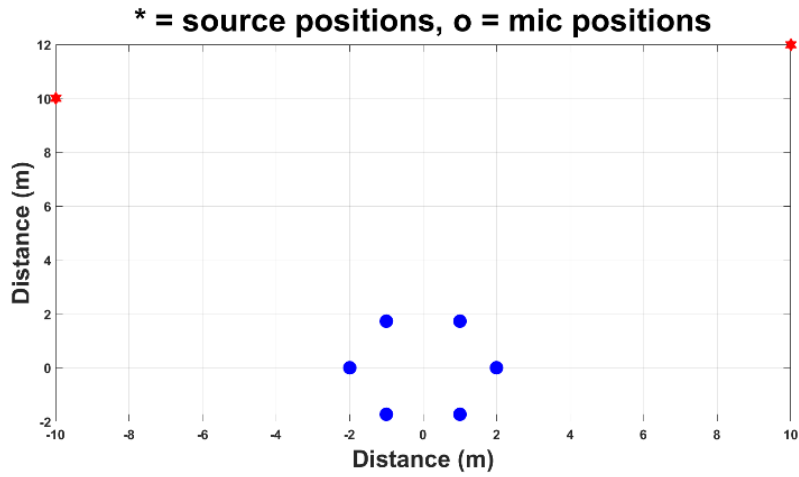


Fig. 6. Direction Detection Steps. The positions for microphone array and sound sources (top);
Two directions detected by steered response power method (bottom).



Fig. 7. Setup process for audio collections using microphone array and off-the-shelf microphone on-site (left), and contact microphone and off-the shelf microphone on-board (middle);
Simultaneous video recording for generating the ground truth data (right).

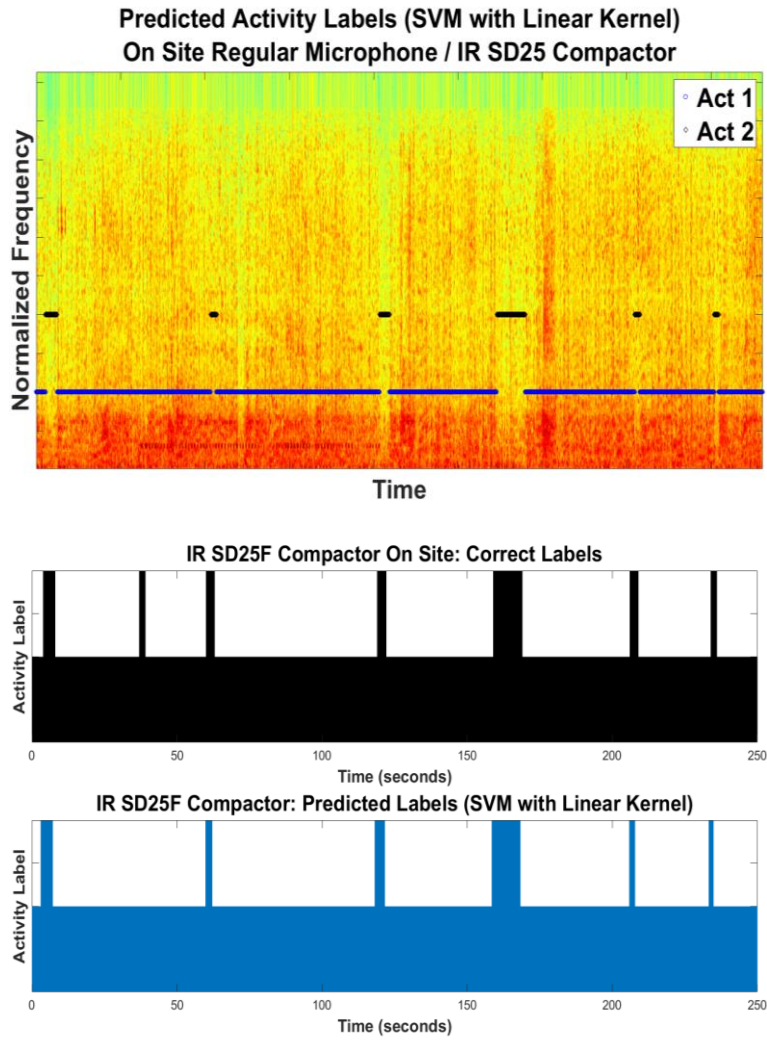


Fig. 8. Results for Ingersoll Rand SD25 Compactor.

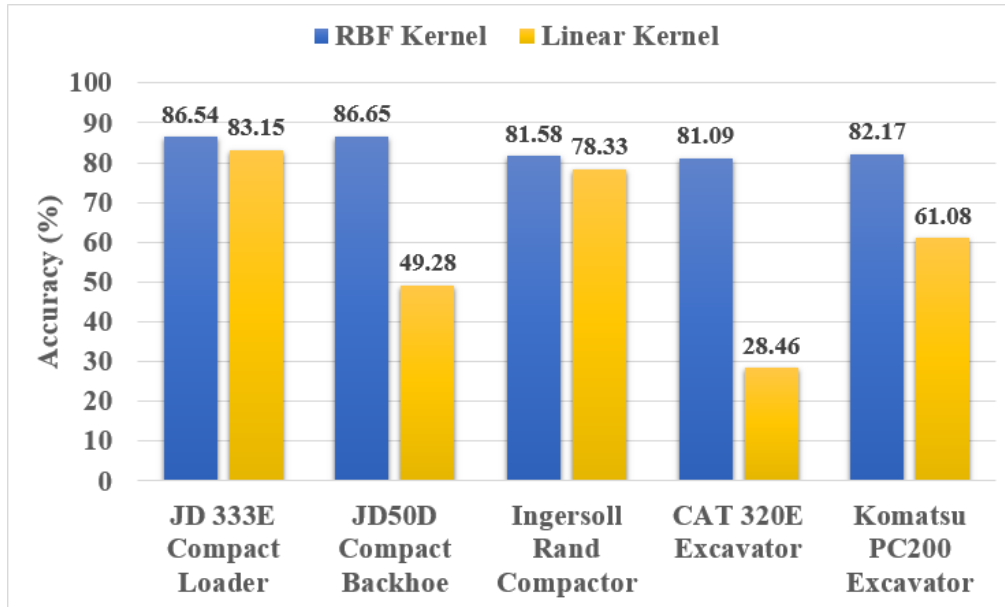


Fig. 9. Summary of the results for implementing RBF and Linear Kernels for activity recognitions of machines (major activities).

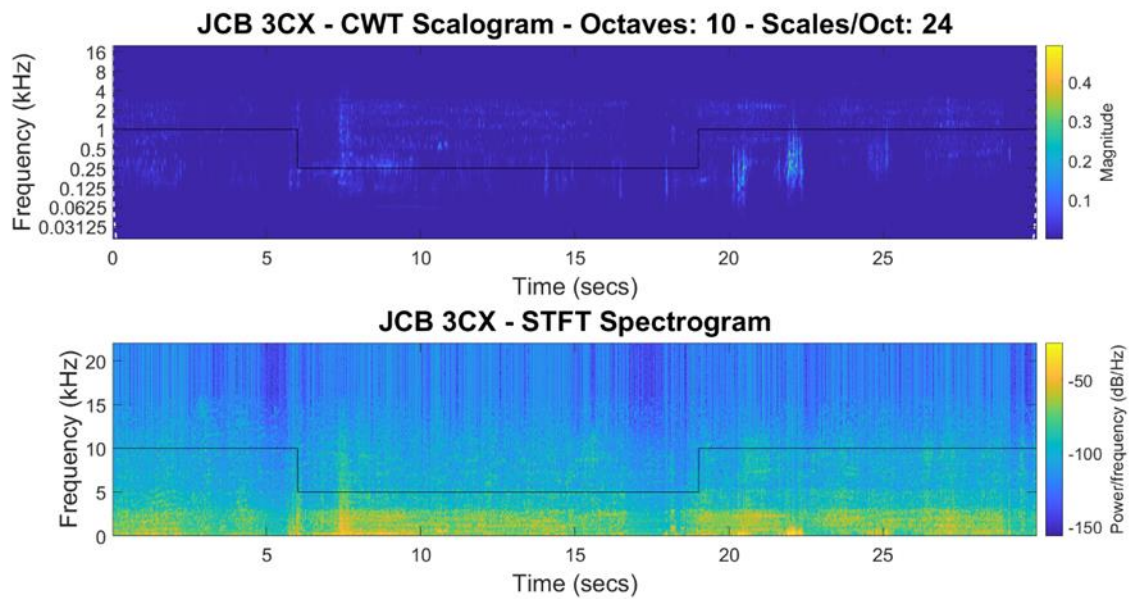


Fig. 10. Sample of comparison results between CWT (10/24) vs. STFT spectrograms-

Machine: JCB 3CX

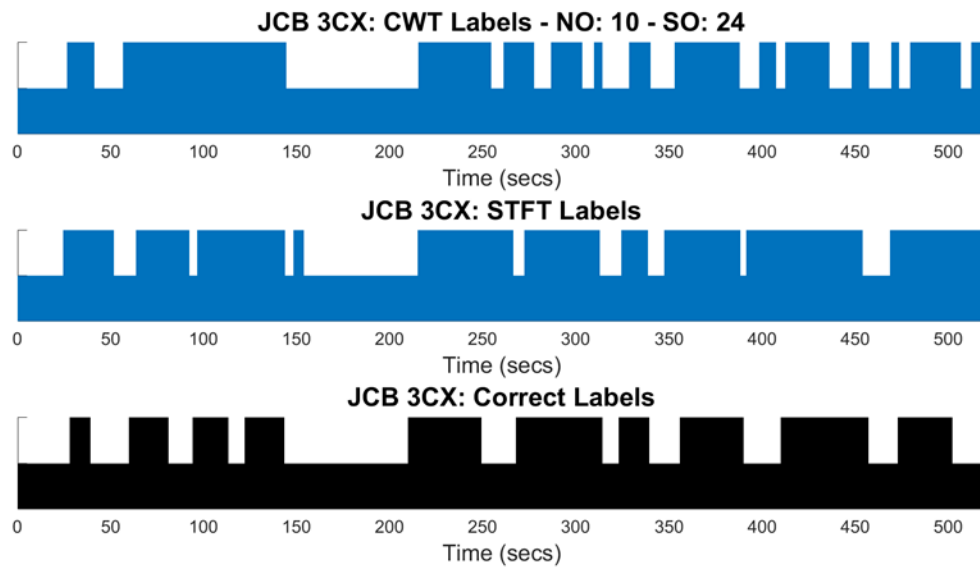


Fig. 11. Labeling comparison between CWT (10/24) vs. STFT spectrograms- Machine:

JCB 3CX