Acoustical Modeling of Construction Jobsites: Hardware and Software Requirements

C. F. Cheng[1], A. Rashidi[2], M. A. Davenport[3], D. V. Anderson[4] and C. A. Sabillon[5]

[1]PhD Student, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA; email: ccheng71@gatech.edu
[2]Visiting Assistant Professor, College of Engineering and Information Technology, Georgia Southern University, Statesboro, GA; email: arashidi@georgiasouthern.edu
[3]Assistant Professor, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA; email: mdav@gatech.edu
[4]Professor, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA; email: anderson@gatech.edu
[5]MS Student, College of Engineering and Information Technology, Georgia Southern University, Statesboro, GA; email: cs10852@georgiasouthern.edu

ABSTRACT

Construction machines and devices often generate distinct sound patterns while performing different tasks, making it possible to extract useful information about jobsites by placing microphones and recording and processing the generated audio files. This paper presents the results of current studies conducted by the authors on the necessary hardware and software for audio modeling of construction jobsites. As the first step, an audio-based system for recognizing activities of construction equipment has been devised. The presented system includes a de-noising algorithm for enhancing the quality of audio files as well as a Short-Time Fourier Transform (STFT) and Support Vector Machines (SVM) for classifying various activities. In the second step, three types of audio recorders (off-the-shelf microphones, contact microphones, and multichannel microphone arrays) and two types of installation settings (microphones mounted on board vs. installed on the job site) have been selected and several experiments were conducted to optimize the hardware settings, tune the algorithmic parameters, and evaluate the different approaches. The results show that for several different types of machines, the accuracy of the audio-based activity recognition system can exceed 85%.

INTRODUCTION

Activity analysis of construction heavy equipment and automatic recognition of various sub-activities (e.g., productive vs. non-productive and idle times) is the first major step toward productivity analysis of a full construction jobsite. In addition to productivity analysis, recognizing various activities taking place by construction machines have several other useful applications including scheduling and costs estimating purposes (Rashidi et al. 2014).

The current states of practice and research for activity recognition of heavy equipment at construction jobsites include implementing active sensors (GPS, MEMS devices such as accelerometers, etc.) and/or passive (processing images/videos using computer vision algorithms) sensors. (Ahn et al. 2015; Brilakis et al. 2011; Golparvar-Fard et al. 2013; Rezazadeh and McCabe, 2012; Rashidi et al. 201; Gong et al. 2011; and Akhavian and Behzadan, 2013).

The authors of this paper have recently initiated a new line of research into the use of audio signals for activity analysis of heavy equipment and thus, acoustical modeling of construction jobsites. The idea is simple: construction machinery and other devices often generate distinct sound patterns while performing their routine tasks at jobsites and it is possible to extract useful information by recording and processing generated audio signals at various locations of a job site. This research aims to bring some of the same tools that have recently been developed in other application areas to bear on the problem of acoustical modeling of construction operations. This builds on recent success in a range of other promising applications, including speech recognition, audio-based navigation of robots, Sound Navigation and Ranging (SONAR) for exploring and mapping oceans and waves (underwater acoustics), and applying ultrasonic signal processing for Condition Based Maintenance (CBM) approaches in manufacturing settings just to name a few (Bengtsson et al. 2004; Greenemeier 2008).

Despite the potential for possible applications, there has been little prior work studying the implementation of audio signal processing techniques in the construction engineering and management domain. In comparison with other active and passive activity recognition methods, audio signals possess the following advantages:

- The majority of 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 paper, 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 activities of multiple machines.
- Computer vision methods are very sensitive to environmental factors such as lighting conditions and occlusions. In addition, 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.
- Compared to video data, audio files have lower data rates and are computationally more efficient.

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

HARDWARE SETTING: DIFFERENT TYPES OF MICROPHONES

Microphones are the primary devices for recording audio. Each microphone contains a surface or moving diaphragm designed to capture electroacoustic waves and generate a corresponding electronic signal. Common approaches for classifying microphones are based on either pickup pattern or type of transducer (Ballou 2015):

- Pickup pattern refers to how the device discriminates between the various directions of incoming sound. Some types of microphones under this classification scheme are omnidirectional microphones, bidirectional microphones, and unidirectional microphones. Pickup pattern is an important factor for acoustical modeling of construction jobsites since in a cluttered site, several pieces of construction devices and machines might work simultaneously at different locations/directions.
- Transducer refers to the device which converts the physical stimulation to an electrical signal. Common types of microphones classified by transducer are carbon microphones, crystal and ceramic microphones, dynamic microphones, condenser microphones, and electret condenser microphones.

Condenser microphones and a variant of so called MEMS (Micro-Electro-Mechanical Systems) microphones are the type of microphone built into the Zoom H1 handy recorder and the

XMOS microphone array board, respectively. According to Yamaha (2016), condenser microphones have good sensitivity to all frequencies, but are highly susceptible to structural vibration and humidity.

In this research, and considering the requirements of construction jobsites, three types of microphones has been selected (Figure 1): 1) an off-the-shelf microphone (Zoom H1 digital handy recorder) 2) Korg CM-200 clip-on contact microphone; and 3) a multichannel microphone array (xCORE-200). A brief description of the contact microphones and microphone arrays is presented here:

- Contact microphones: Contact microphones are built with a unidirectional, piezoelectric transducer, which is designed to be less susceptible to air-carried sound waves. The Korg CM-200 microphone, selected for data collection purposes in this research, is commonly used in applications that involve capturing sound from a particular music instrument when recording or practicing with an entire music band. Contact microphones have great potential for on-board applicability within heavy construction equipment with the advantage of not being liable to structural vibrations, as opposed to condenser microphones (Figure 2).

- Microphone arrays: A microphone array is constituted by a group of two or more microphones working in tandem, commonly arranged in linear, rectangular, and circular patterns. These microphones are usually omnidirectional; however, some microphone arrays are built completely by directional microphones or a combination of omnidirectional and directional microphones (Brandstein and Ward 2001). For this project, an XMOS xCORE-200 has been chosen. The applications for this type of array-based microphones are based on beamforming techniques and include: speech enhancement, speech recognition, source localization, noise reduction, echo cancellation, and separation of acoustic signals.

The XMOS xCORE-200 microphone array board, is equipped with seven omnidirectional MEMS microphones with pulse density modulation (PDM) output, 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 depicted in Figure 2. The seven microphones built into this board are targeted, but are not strictly limited to, Voice User Interface (VUI) applications. As shown in Figure 1, one microphone is placed at the center of the board and the remaining six are distributed equidistant around the board edge.

Figure 1: Different types of microphones used in the project: Zoom H1 digital handy recorder mounted on a tripod (left); Korg CM-200 clip-on contact microphone attached to CAT loader (middle); xCORE-200 microphone array evaluation board – top (left).

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

In order to assess the hardware and software requirements for acoustical modeling of construction jobsites, a basic audio-based model for activity analysis of construction heavy equipment has been devised by the authors. This algorithm consists of the following major components and algorithms (Figure 2):

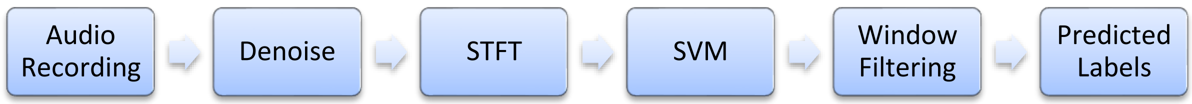

Figure 2: The audio-based model for activity analysis of construction heavy equipment.

- A denoising algorithm is used to reduce the background noise and enhancing quality of the audio signal. In this research, the denoising algorithm proposed by Rangachari and Loizou (2006) has been selected and implemented.

- A short-time Fourier transform (STFT) is used to covert the audio signal into a time-frequency representation.

- Support Vector Machines (SVM) have been implemented as the major machine learning tool for training and testing the system for identifying different activities within each captured audio signal. To achieve this goal, the authors used the LIBSVM MATLAB package. It is well known that the performance of the SVM algorithm highly depends on the selected kernel function (Rashidi et al. 2016). In this research, the two more common kernel functions (linear and radial basis function) has been selected and experiments have been conducted using each kernel separately.

- A window filtering approach is used to label time frames of different activities. 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.

More detailed information about the implemented audio-based system can be found at (Cheng et al, 2016). Finally, it is important to mention that in this paper, we only focus on activity analysis of single machines. The more complicated (and realistic) case of dealing with multiple machines simultaneously is not within the scope of the current work and will be investigated by the authors in the future.

EXPERIMENTAL SETUP AND PRELIMINARY RESULTS

In order to evaluate the performance of the audio base activity analysis system under various hardware and software settings, a number of construction machines from different jobsites have been selected. Audio data sets from individual pieces of equipment performing routing actions were captured using four recording devices simultaneously (Figure 3): one Zoom H1 digital handy recorder was placed on a tripod on site; one XMOS xCORE-200 USB microphone array board connected to a laptop computer on site; another Zoom H1 handy recorder placed on board the heavy equipment, usually inside the cup holder; and the Korg CM-200 contact microphone clamped to a flat surface inside the cabin of the construction 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 audio was recorded through Audacity®, an open source software.

In addition to audio recording, a video sample was taken using a regular cell phone camera to serve as a reference for manual action classification into major activities (e.g., digging, loading, dumping, crushing rock) and minor activities (e.g., swinging, maneuvering, extending arm). Once all recording devices were running, an air horn was used to generate a loud noise intended to be captured on all recordings. That would be the reference signal for synchronizing all data.

Figure 4 illustrates the generated results for one sample machine, A CAT 320D Backhoe excavator, using a Zoom H1 digital handy recorder placed relatively close to the machine on the jobsite. The presented results are based on implementing SVM with linear kernel. The top part of the figure presents the normalized frequency for the machine's audio signal, while the middle and bottom parts show the actual (blue) versus predicted (black) 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 CAT backhoe excavator.

Figure 3: Setup process for audio collections using microphone array (left) and contact microphone (middle); Simultaneous video recording for generating the ground truth data (right).

Details of comparison results for different machines and under different hardware and software settings have been summarized in Tables 1-3 and Figure 5. By careful observation of these results, the following conclusion marks can be achieved:

1- For on board settings, using contact microphones generates slightly better results and thus, is a better option (Table 1).
2- The current case studies show that placing regular microphones on jobsites generates better results compared to placing contact microphones in the cabin. One reason for this phenomenon is that the contact microphones directly mounted on board would be affected by engine noise and vibrations; however, further investigations using several other case studies and under various jobsite conditions might be required to fully substantiate this conclusion (Tables 2 and 3).
3- In most cases, especially for recognizing major activities, the radian basis function kernel outperforms the linear kernel (Figure 5).
4- There is no significant difference between using regular microphones and microphone arrays for recognizing activities of single machines. The main reason is that there is only one source, and thus only one direction for the generated audio signals. It is anticipated that there will be a significant difference between the performance of the regular microphones and microphone

arrays in the case of existence of multiple machines (or multiple sources of generating audio). Microphone arrays will be able to detect multiple audio directions from multiple audio sources. The authors will investigate this condition (multiple machines working simultaneously in the job site) in the future.

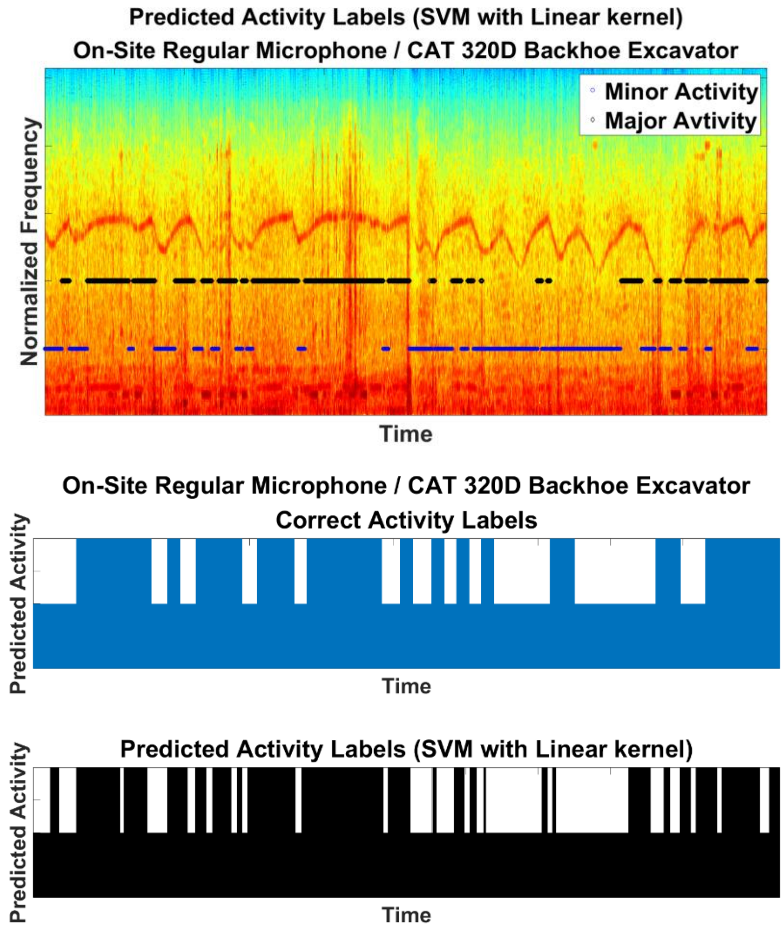

Figure 4: Results for CAT 320D backhoe excavator

Table 1: Comparison between contact and regular microphones (both on board/ SVM with 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%

Table 2: Comparison between contact microphone (on board) and regular microphone (on site).

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%

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%

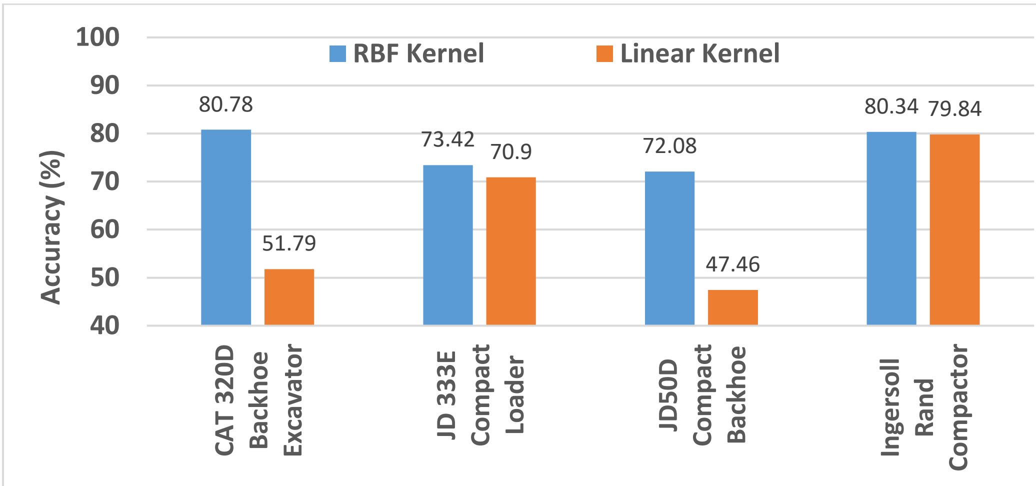

Figure 5: Summary of the results for implementing RBF and Linear Kernels for activity recognitions of machines (major activities).

SUMMARY AND CONCLUSION

This paper presented an innovative audio-based method for activity analysis of heavy equipment and acoustical modeling of construction jobsites. A basic audio based activity recognition model has been developed for single machines and has been tested under various hardware and software settings. The experimental settings include applying three major types of microphones (off-the-shelf microphones, contact microphones; and microphone arrays) and two placement settings (on board vs. on site). The preliminary comparison results have been presented in the previous sections. In addition, the audio-based model has been tested using two major types of kernels (linear and RBF) and the results indicated the better performance of the RBF kernel in most cases. As the extension of the current research, the authors plan to work on the following items in future:

- Implementing and testing the proposed audio-based model using more data sets and under various conditions
- Evaluation and testing the required hardware and software settings for multiple machines working at the construction jobsites
- Developing more robust algorithms to recognize construction activities into more detailed items (splitting minor and major activities into sub-activities)

ACKNOWLEDGEMENT

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

Akhavian, R., and Behzadan, A. (2013). "Knowledge-based simulation modeling of construction fleet operations using multimodal-process data mining." *Journal of Construction Engineering and Management*, 139(11): 04013021.

Ahn, C., Lee, S., and Peña-Mora, F. (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.

Bengtsson, M., Olsson, E., Funk, P., and Jackson, M. (2004). "Technical design of condition based maintenance system—A case study using sound analysis and case-based reasoning." *In Proc. International Conference of Maintenance and Reliability (MARCON)*, Knoxville, TN, May 2004.

Brilakis, I., Fathi, H., and Rashidi, A. (2011). "Progressive 3D reconstruction of infrastructure with videogrammetry." *Automation in Construction*, 20(7): 884-895.

Ballou, Glen. Handbook for Sound Engineers. 5th. New York: Focal Press, 2015.

Cheng, C., F., Rashidi, A., Davenport, M. and Anderson, D. "Audio Signal Processing for Activity Recognition of Construction Heavy Equipment." *In Proc. 33rd International Symposium on Automation and Robotics in Construction (ISARC 2016)*, July 18-21, 2016, Auburn, AL.

Golparvar-Fard, M., Heydarian, A., and Niebles, J.C. (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., Caldas, C.H., and Gordon, C. (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. "A positioning system that goes where GPS can't." January 23, 2008. Scientific American, http://bit.ly/2hsVPNJ.

National Oceanic and Atmospheric Administration. "What is Sonar?" June 1, 2016. http://oceanservice.noaa.gov/facts/sonar.html.

Rangachari, S., and Loizou, P. (2006). "A noise estimation algorithm for highly nonstationary environments." *Speech Communication*, 28: 220-231.

Rashidi, A., Fathi, H., and Brilakis, I. (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., Nejad, H.R., and Maghiar, M. (2014) "Productivity estimation of bulldozers using generalized linear mixed models." *KSCE Journal of Civil Engineering*, 18(6): 1580-1589.

Rashidi, A., Sigari, M.H., Maghiar, M. and Citrin, D. (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 McCabe, B. (2012). "Part based model and spatial–temporal reasoning to recognize hydraulic excavators in construction images and videos." *Automation in Construction*, 24: 194–202.

XMOS Ltd. "xCORE Microphone Array Hardware Manual." March 2, 2016.

Yamaha. "Which types of Microphone Are Used with PA systems?" April 2, 2016. http://bit.ly/2cMoi0Y.