Fall 2017

Audio-Based Productivity Forecasting of Construction Cyclic Activities

Chris A. Sabillon

Recommended Citation

AUDIO-BASED PRODUCTIVITY FORECASTING OF CONSTRUCTION CYCLIC
ACTIVITIES

by

CHRIS ANDRES SABILLON

(Under the Direction of Biswanath Samanta)

ABSTRACT

Due to its high cost, project managers must be able to monitor the performance of construction heavy equipment promptly. This cannot be achieved through traditional management techniques, which are based on direct observation or on estimations from historical data. Some manufacturers have started to integrate their proprietary technologies, but construction contractors are unlikely to have a fleet of entirely new and single manufacturer equipment for this to represent a solution. Third party automated approaches include the use of active sensors such as accelerometers and gyroscopes, passive technologies such as computer vision and image processing, and audio signal processing. Hitherto, most studies with these technologies have aimed to activity identification or to identifying active and idle times. Given that most actions performed with construction machinery involve cyclic activities, cycle time estimation is much more relevant. In this study, hardware and software requirements were optimized toward that goal. This approach had three facets: first, signal spectral analysis was performed through the short-time Fourier transform (STFT) and the continuous wavelet transform (CWT) for comparison; second, audio and active sensor data have been submitted to a machine learning framework for activity classification accuracy comparison; and, third, Bayesian statistical models were used to include historical data for cycle time estimation enhancement. As a result, audio signals have been used along with a Markov-chain-based filter to achieve cycle time estimation with an accuracy of over 81% for up to five days of single-machine operation.

INDEX WORDS: Markov models, Bayesian statistics, Machine learning, Support vector machines, Continuous wavelet transform, MEMS accelerometers, Audio signal processing, Productivity, Construction heavy equipment

AUDIO-BASED PRODUCTIVITY FORECASTING OF CONSTRUCTION CYCLIC

ACTIVITIES

by

CHRIS ANDRES SABILLON

B.S., Universidad Tecnológica Centroamericana, Honduras, 2012

A Thesis Submitted to the Graduate Faculty of Georgia Southern University in

Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

STATESBORO, GEORGIA

AUDIO-BASED PRODUCTIVITY FORECASTING OF CONSTRUCTION CYCLIC
ACTIVITIES

by

CHRIS ANDRES SABILLON

Major Professor: Biswanath Samanta
Committee: Abbas Rashidi
                     Minchul Shin

Electronic Version Approved:

December 2017

DEDICATION

To the Sabillons, Telle, Honduras, and my professors for their advice.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

CWT	Continuous Wavelet Transform
MEMS	Micro-Electro-Mechanical-Systems
RBF	Radial Basis Function
STFT	Short-Time Fourier Transform
SVM	Support Vector Machine

CHAPTER 1

INTRODUCTION

1.1   Problem Statement

Research shows that non-value adding activities consume between 50% and 75% of the total time spent on a construction jobsite, as opposed to less than 10% inactive time in manufacturing (Construction Industry Institute 2014). A key cause to this situation is that manufacturing activities are constantly measured and controlled through state-of-the-art automated systems, while construction management is much more rudimentary. In fact, state-of-practice performance monitoring is based on manual data collection performed through direct observation of live activities or video streams. This activity is time consuming, expensive in terms of labor cost, prone to human error, and rarely allows for immediate application of corrective measures.

A major restriction to developing an automated performance monitoring system for the construction environment is that projects are greatly diverse and that the activities within them are hard to classify. Even projects with the same design are different due to unique factors, such as: soil conditions, weather conditions, accessibility to power sources and utilities, logistics for supply and access, regulatory requirements, contractual constraints, personnel skill, and management level of expertise (Intergraph Corporation 2012). These conditions complicate not only live monitoring of construction sites but also project planning.

Project managers commonly allocate most of a project's cost toward owning and operating heavy equipment. Thus, equipment type and quantity must be selected optimally to obtain an acceptable financial yield. To do so, an expected production rate per piece of machinery is calculated through historic data, manufacturer manuals, or guides (Peurifoy, et al. 2010, Caterpillar

2017). Projected productivity rarely matches actual productivity due to the same factors that make each project unique, so managers attempt to use common-practice monitoring to apply corrective measures. Statistics show that this combination is ineffective. In fact, the need for an automated monitoring system has been long identified by the industry and academia (Akhavian and Behzadan 2014, Ahn, Lee and Peña-Mora 2012, Khosrowpour, Nieblesb and Golparvar-Fard 2014, Navon 2005, Tajeen and Zhu 2014, Teizer, et al. 2010). With an automated performance monitoring system for heavy equipment, managers would be able to apply on-time corrective measures that would not only reduce idle times but also help avoid on-site struck-by accidents.

1.3   Hypothesis

If an activity recognition and cycle time estimation framework is developed based on jobsite sensor data recordings, then it can be used to as a foundation toward a real-time construction equipment monitoring system with universal compatibility.

1.2   Scope of Present Work

For earthmoving operations, yield is calculated in terms of volume of displaced material or finished surface area. Tractors, loaders, excavators, and graders are the principal machinery used to execute these tasks. Productivity estimations for all these types of equipment are inversely proportional to cycle time. Other parameters for calculation like bucket capacity, fill factor, and blade size, are fairly constant because they depend on equipment design and type of material being worked with. Thus, an attempt for real-time monitoring of construction equipment must focus on calculating cycle times accurately.

Some manufacturers have started to integrate their own monitoring technologies, but contractors are unlikely to have a fleet of entirely new and/or single manufacturer equipment for

this to be solution. Third party approaches toward an automated heavy equipment monitoring system are three-fold: the use of active sensors such as GPS, accelerometers, and gyroscopes (Ahn, Lee and Peña-Mora 2014, Teizer, et al. 2010, Torrent and Caldas 2009); the use of passive technologies such as computer vision and image processing (Bügler, et al. 2014, Golparvar-Fard, Heydarian and Niebles 2013, Gong, Caldas and Gordon 2011, Zhu, et al. 2016); and audio signal processing for activity recognition (Cheng, et al. 2017, Cho, Lee and Zhang 2017).

Most studies focus on activity recognition or comparing active vs. idle times, not cycle time estimation, which would be more relevant. Additionally, no comparison among these technologies has been made. Regarding audio signal processing for construction equipment activity recognition, various microphone types and placement settings have been compared and machine learning algorithms have been enhanced (Cheng, Rashidi, et al. 2017). Consequently, it was determined that on site microphone placement produced consistently better results than microphones placed on board the equipment and that the radial basis function (RBF) kernel selection for the support vector machine (SVM) classifier produced better results than the linear kernel. The present work aimed to further optimize hardware and software configurations via the following key objectives:

- Examine options for audio time-frequency feature extraction by comparing results obtained using the short-time Fourier transform (STFT) versus results obtained using the continuous wavelet transform (CWT).

- Evaluate construction equipment activity classification accuracy by comparing two major monitoring approaches: active sensors and audio signal processing

- Evaluate potential for activity labeling classification accuracy improvement by combining audio data and active sensor data as input.

- Implement Bayesian methods to develop a cycle time forecasting system capable of being implemented over multiple days of operation.

1.4   Research Limitations

It is important to note that this research was performed under the following conditions:

- This study was performed using jobsite sensor recordings. Thus, the performance monitoring framework is not applied in real-time conditions.

- Audio and active sensor data was taken for single machines working independently. More realistic jobsite conditions involve multiple machines working simultaneously.

- Computational processing and memory capabilities (7th Generation Intel Core i7 laptop with 16 GB of RAM) proved to be a limiting factor while performing certain algorithms, like the wavelet transform.

- Active sensor data collection was performed through the MATLAB Support Package for Android Devices, which requires the mobile device, mounted on board the heavy machinery, to be connected via Wi-Fi to a host computer on site. Interruptions to wireless connectivity hindered the possibility of obtaining a representative amount of field data.

1.4   Organization of Thesis

The rest of the thesis is organized in the following order:

Chapter 2 is the literature review divided in six sections. First, a detailed introductory overview about the need for a monitoring system and state-of-practice techniques is presented. In the

following three sections, the main third party automated approaches for construction equipment monitoring (active sensors, passive sensors, and audio signal processing) are described, along with advantages and limitations of each. In the fifth section, the Fourier transform, and the wavelet transform are discussed thoroughly. Finally, Bayesian models and applications of such in the manufacturing and construction industries are detailed.

Chapter 3 covers the research methodology for this study, which is subdivided in three parts. First, the audio signal processing framework is presented, including data collection setup and result evaluation criteria. Second, the outline for cycle time estimation using audio signals as input and Bayesian models for filtering is described. Finally, the active sensor data processing framework is described. It is important to note that cycle time estimation using active sensors was not part of the scope of this thesis. Active sensors were only used to label activities and compare classification accuracy against the audio framework.

Chapter 4 presents the experimental results and related discussions. First, results obtained by processing audio signals through the STFT versus the CWT are compared. Second, active sensor (accelerometers) data labeling accuracy is compared versus audio signal processing labeling accuracy. Third, a brief evaluation of the potential benefit of combining audio and active sensor data is executed. Finally, optimal hardware and software settings are used for cycle time estimation.

Chapter 5 presents a summary of the present work along with recommendations for future work.

CHAPTER 2

LITERATURE REVIEW

2.1   Automation for Construction Site Monitoring

Most non-farming labor efficiency has at least doubled since the 1960s, except for construction industry. In fact, the Lean Construction Institute (2017) estimates that over 70% of projects are over budget and delivered late and affirms that the construction industry records about 800 deaths and thousands of injuries per year. A real-time, automated performance monitoring system that may allow the construction industry to reduce waste, improve safety, and manage labor efficiently is a pressing need considering that construction industry contributes to at least 10% of the gross national product (Navon 2005). One of the major impediments for developing such automated performance monitoring system for the construction environment is that projects are diverse and that the activities within them are hard to classify, even during different stages of the same project.

Construction heavy equipment is a key component in any jobsite and represents a high potion of total project costs. Thus, measuring and analyzing its operation is essential for productivity improvement, not only to control current projects, but also to update historical databases (Ahn, Lee and Peña-Mora 2014). These databases allow for better planning in future projects that, in consequence, can yield a safer working environment and reduce carbon footprint.

Hitherto, construction equipment is commonly monitored through a sampling concept that relies on an operator manually filling documentation to record what is taking place on the field. This method relies on the idea that the time spent on value-adding activities is an indirect measurement of equipment productivity (Rashidi 2015). It is evident that this approach is not ideal because it is subject to human error, it is time-consuming, and does not allow immediate response

and corrective measures by project managers (Rezazadeh, Dickinson and McCabe 2013). That is the reason why an automated, real-time monitoring and tracking system is an urgent need.

Some manufacturers have started to integrate their proprietary monitoring devices to new equipment, but construction contractors are unlikely to have a fleet of entirely new and single manufacturer equipment for this to represent a solution. Third party attempts to achieve an automated heavy equipment monitoring and tracking system are based either in using active sensors or in using passive sensors. Active sensors are those that produce a change in current or voltage due to an external environmental stimulation. Passive sensors are those that require an external source of excitation because they produce a passive signal, e.g., change in resistance or change in capacitance. These technologies will be discussed thoroughly in the following sections.

2.2   Active Sensors: Gyroscopes and Accelerometers

The basic principle behind active sensor technology consists in mounting gyroscopes and accelerometers onto construction equipment to detect movement in the three-dimensional Cartesian coordinate system. Specific movement patterns are then related to a specific activity. These activity-related patterns are trained to a computer through a supervised machine learning algorithm. Once several activities have been trained, a library has been created. This library is, finally, validated and used for real-time activity recognition.

Promising advances in active sensor application for activity recognition of construction equipment have been achieved in a study performed by Ahn, Lee, and Peña-Mora (2014). They successfully used low-cost micro-electro-mechanical-systems (MEMS) accelerometers to identify three modes of excavator operation (engine-off, idling, and working) with an accuracy of up to 93%. Nonetheless, these results have been obtained in a controlled environment, with one specific

machine, and with a limited quantity of operations. Additionally, MEMS have to be mounted on the equipment, which may represent a setback while using this system on leased or rented machines.

Other significant results were obtained through the extraction of multi-modal data from integrated cell phone sensors, specifically: GPS, gyroscope, accelerometer (Akhavian and Behzadan 2014). Results from implementation of several machine learning classifiers (k-nearest neighbors [K-NN], decision tree, logistic regression, support vector machines [SVM], and neural networks) showed successful cataloging of engine on, engine off, idling, and maneuvering activities. However, all tested classifiers were inaccurate in other activities that involved beam and bucket movements. This can be explained by the fact that cell phones were mounted in cabin, where the patterns generated by beam and bucket movements were attenuated by the distance and by the movement patterns produced due to other classified activities.

2.3   Passive Sensors: Computer Vision

The evolution of computational capacities, communication networks, and high-resolution, digital cameras has presented computer vision as an interesting instrument for unlimited applications. Furthermore, smart devices put this all these capabilities on hands of any individual. A worker in a construction jobsite can carry an inexpensive camera or smartphone, record video, and provide potential means for productivity estimation through computer vision algorithms. A real-time system could also be executed by placing recording devices on a jobsite and applying a computer vision algorithm. The principle of operation for computer vision consists of four basic steps depicted in Figure 2.1: equipment recognition, equipment tracking, action recognition, and performance assessment.

The equipment recognition step is currently based on one of three techniques: part-based recognition, appearance-based recognition, and feature-based recognition. Once a specific equipment is recognized, the equipment is tracked to reduce the viewing area and avoid noise from background dynamics. Finally, action recognition is performed to be able to achieve a performance assessment. It is important to note that an action is composed of several activities and movements. For example, the action digging a foundation consists of activities like digging, swinging, and dumping, which consist of several movements like raising the arm, lowering the arm, swinging the bucket, grabbing soil, and pushing soil (Cheng, et al. 2016). That definition for action is one of the pillars of this study.
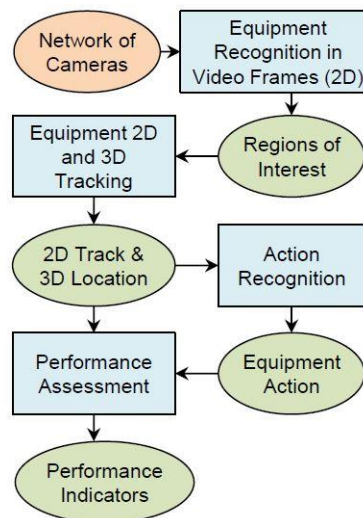
Figure 2.1: Flowchart for computer vision-based performance monitoring of heavy equipment (Cheng, et al. 2016).

Numerous approaches have been documented for vision-based action recognition of heavy equipment and workers. Kim and Caldas (2013) proposed an action recognition method that related workers and their interactions with specific objects or tools as a work rate measurement system for productivity estimation. Azar, Dickinson, and McCabe (Rezazadeh, Dickinson and McCabe 2013) developed a hybrid framework that involved object recognition, tracking, and action recognition of dump trucks and loaders to estimate loading cycles. Golparvar-Fard,

Heydarian, and Niebles (2013) used a multi-class support vector machine (SVM) classifier to recognize and localize equipment actions given a video sequence previously recorded from a fixed camera. Experimental results for this SVM classifier yielded average accuracies of 86.33% and 98.33% for an excavator and a dump truck, respectively. Bügler, et al. (2014) proposed a method for tracking the progress of earthmoving actions by combining two vision-based technologies: photogrammetry and video analysis. Photogrammetry was used to determine the total volume of removed soil at given intervals, while video analysis was used to determine equipment active and idle times. Combining the data obtained from both sources served them to calculate individual equipment productivity and specific site performance factors.
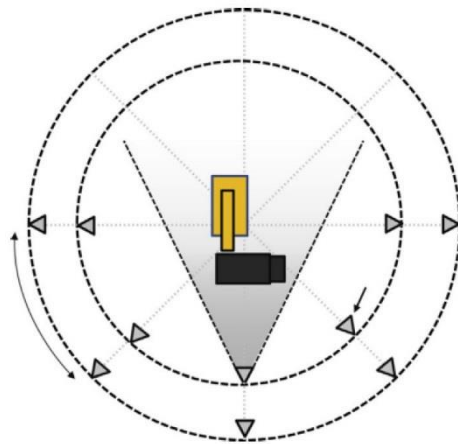
Figure 2.2: Layout for computer vision-based equipment action recognition (Golparvar-Fard, Heydarian and Niebles 2013).

Although active sensor action recognition and computer vision-based action recognition have a promising panorama, there is still opportunity for other technologies. The use of active sensors is still in its early stages of development, while computer vision-based technologies have several setbacks. Cameras require a proper level of illumination in order to capture quality video. Complete darkness or direct sunlight are inevitable sources of noise in projects that are active without interruption during sunny days, cloudy days, rainy days, or even nights. Furthermore, cameras have a limited field of view. The necessity to place cameras in a circular pattern in order

to cover all angles in a jobsite is depicted in Figure 2.2. This type camera network highly increments the need of processing power and significantly hinders the possibility of generating low-cost, real time framework for activity recognition. Finally, most construction projects are performed by contractors and subcontractors that usually have a problem with being recorded continuously due to privacy issues or legal issues that could arise from a job-related accident.

2.4   Audio Signal Processing

2.4.1   Microphones

Microphones are the primary devices used for audio recording. A microphone generally contains a moving diaphragm or surface designed to capture electroacoustic waves and generate a corresponding electrical signal. Sound sources have different characteristics (e.g., waveform, phase, dynamic range, attack time, frequency), so microphones are designed specifically depending on the application. Pickup pattern or type of transducer are common characteristics considered for microphone classification (Ballou 2015).

A microphone is designed to capture different directions of incoming sound with different intensity. This is the microphone's pickup pattern. Some common microphone designs under the pickup pattern classification scheme are omnidirectional, bidirectional, and unidirectional or cardioid microphones. In an omnidirectional microphone, pickup pattern is equal in all directions. Omnidirectional microphones are particularly useful in a setting that requires all audio elements to be captured, as in an orchestra. In a bidirectional microphone, pickup pattern is equal on opposite directions and negligible 90° from these. Bidirectional microphones are particularly useful when the audio sources of interest are placed in front of each other, like two speakers holding a conversation face to face. In a unidirectional microphone, the pickup pattern has a cardioid shape

facing to one direction. Unidirectional microphones and other variations of directional microphones are the most widely used of all previously mentioned microphones because they allow to focus on one specific audio source. More details about these and other microphones are provided in Figure 2.3.

Microphone	Omnidirectional	Bidirectional	Directional	Supercardioid	Hypercardioid
Directional response characteristics					
Voltage output	$E = E_o$	$E = E_o \cos\theta$	$E = \frac{E_o}{2}(1+\cos\theta)$	$E = \frac{E_o}{2}\left[(\sqrt{3}-1)+(3\sqrt{3})\cos\theta\right]$	$E = \frac{E_o}{4}(1+3\cos\theta)$
Random energy efficiency (%)	100	33	33	27	25
$\dfrac{\text{Front response}}{\text{Back response}}$	1	1	∞	3.8	2
$\dfrac{\text{Front random response}}{\text{Total random response}}$	0.5	0.5	0.67	0.93	0.87
$\dfrac{\text{Front random response}}{\text{Back random response}}$	1	1	7	14	7
Equivalent distance	1	1.7	1.7	1.9	2
Pickup angle (2θ) for 3 dB attenuation	-	90^0	130^0	116^0	100^0
Pickup angle (2θ) for 6 dB attenuation	-	120^0	180^0	156^0	140^0

Figure 2.3: Performance characteristics of several microphones (Ballou 2015).

The transducer is the device that converts a physical stimulus into an electrical signal output. Common microphones regarding the type of transducer outline are carbon, crystal, and ceramic microphones, condenser microphones, dynamic microphones, and electret microphones. The transducer is a determining factor when it comes to microphone's response to sound frequencies, structural vibrations, temperature, humidity, and other environmental factors.

A multifaceted jobsite is likely to involve multiple pieces of equipment that may work simultaneously generating sound from various directions and environmental conditions that vary along the year or depending on the activities being executed. Pickup pattern must be flexible enough to allow focusing on a source on interest and the transducer must be robust enough to withstand inclement weather and other contingencies while maintaining stable audio signal recording characteristics. Thus, properly selecting microphone type and placement setting is a key

for acoustical modeling of construction jobsites. In a separate work, hardware requirements have been analyzed thoroughly (Cheng, Rashidi, et al. 2017) and the XMOS xCORE-200 multichannel array microphone has been selected accordingly. This microphone array board consists of is hardware and reference software platform equipped with seven omnidirectional MEMS transducers with pulse density modulation (PDM) output. As illustrated in Figure 2.4, one microphone is located on the center of the board and the remaining six are distributed equidistantly on a circular pattern along the edge of the board. MEMS microphones are a variant of condenser microphones, which have good sensitivity to all frequencies, but are highly susceptible to structural vibration and humidity (Yamaha 2016). Liabilities that can be easily overcome during controlled data collections but must be kept in mind for a permanent application.

Figure 2.4: xCORE-200 microphone array evaluation board – top (XMOS Ltd. 2016).

2.4.1   Advantages of Audio Signal Processing

Audio signal processing for action recognition can certainly signify several advantages over other passive and active sensor technologies. According to Rashidi (2015) these advantages are,

at least, four: first, some complex actions are easier to recognize through sound (e.g., a hydraulic hammer attached to an excavator arm produces insignificant movement to be identified through accelerometers or cameras, but does produce a characteristic sound); second, characteristic equipment sound is independent of the operator; third, microphones can be placed on a jobsite and record data in an omnidirectional manner, whereas, accelerometers must be fixed onto the equipment and cameras have a limited field of view; fourth, data transmission rates for audio signals are usually 400 times smaller than video signals, which considerably reduces processing requirements.

This set of advantages has resulted in the application of audio signal processing in fields such as medicine, industrial automation, robotics, identification and tracking, gadgets, and military technology. Commonly known applications include: ultrasonic imaging for obstetrics, tissue scanning, and engineering structural analysis; ultrasonic sensors for object detection and distance measurement in industrial automation and robotics; sound navigation and ranging (SONAR) for navigation and tracking; and sound detection and ranging (SODAR) for meteorology and wind energy feasibility analysis. The following sections; however, are focused on providing insight on applications that involve action recognition.

2.4.1 Audio Signal Processing in Medicine

Several techniques have been applied to monitor patients with chronic disease such as asthma and chronic obstructive pulmonary disease (COPD). Specifically, to evaluate the adherence to inhaler medication, which includes taking doses in a consistent schedule and method. The inhaler compliance assessment (INCA) device, depicted in Figure 2.5 (left), was developed as an approach to evaluate inhaler adherence. This device is attached to a widely used variety of inhalers. When the patient opens the mouthpiece to take a dose, the INCA device takes an audio recording with a

timestamp. Figure 2.5 (right) is an example of a typical audio sample after applying the fast Fourier transform (FFT) algorithm to extract its frequency features. 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 in a large case study. 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).
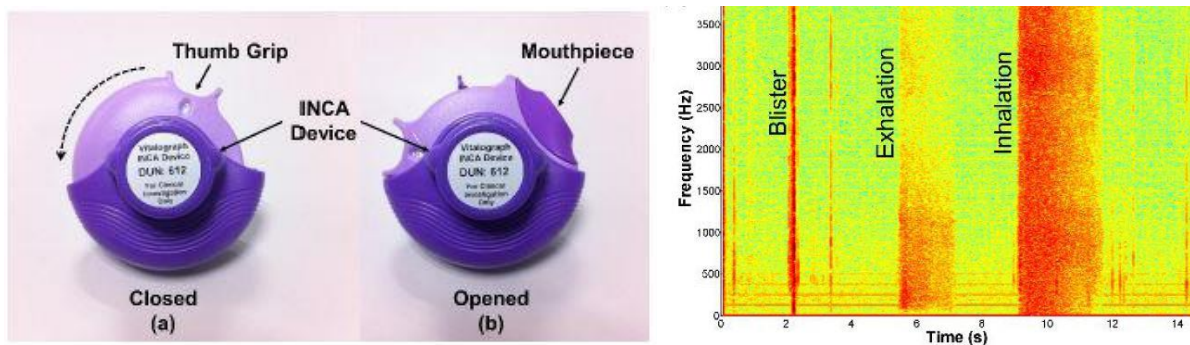
Figure 2.5: Inhalator with INCA device (left). Spectrogram of one sample taken by INCA device (right) (Holmes, et al. 2014).

2.4.2   Audio Signal Processing in Manufacturing and Power Industry

Audio signal processing, specifically ultrasound, has important applications as Condition Based Maintenance (CBM). Ultrasonic sound is usually in the range of 20 kHz to 100 kHz, which is far beyond the human ear audible range. Therefore, certain techniques are necessary to interpret ultrasound. The two major approaches to the use of ultrasound in CBM are direct interpretation by a trained technician and computer-based analysis. Direct interpretation by a trained technician is performed through a technique of heterodyning or translating ultrasound to an audible range that can be interpreted through headphones and a decibel display. Computer analysis is performed via a software that records several audio benchmarks of a failure cases and proper functioning. Once a case library is created, future recordings are compared to the audio benchmarks to produce a

condition diagnosis. Some typical applications of ultrasound processing in 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).

Bengtsson, et al. (2004) define computer-based analysis as a three-module process, as depicted in Figure 2.6. 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 information and feature extraction to identify characteristic features of the audio sample and create a vector. Once a vector is created, the condition monitoring and diagnosis module consists in comparing the audio sample to an existing library and provide a condition diagnosis. If the software is in a learning process, when a new vector is identified it is stored to the case library. This model is characteristic to most audio-based machine learning practices. Thus, it can be considered as a basis for this study.
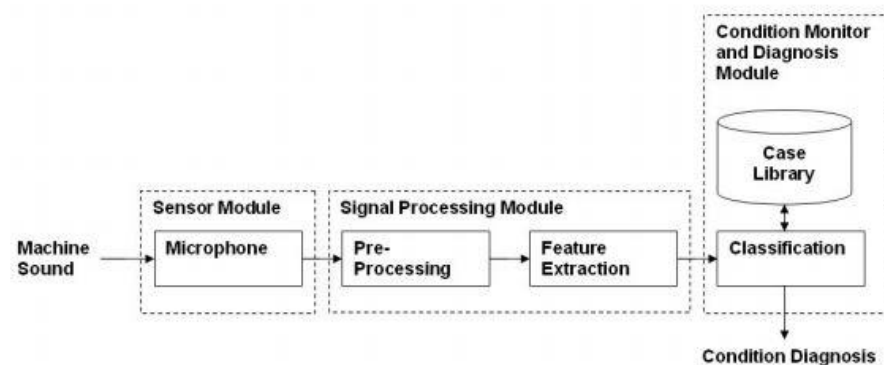
Figure 2.6: Schematic of the case-based fault diagnosis system with its three steps to condition diagnosis (Bengtsson, et al. 2004).

2.5   Frequency Domain Analysis

It is generally more relevant to analyze a signal in terms of its frequency components. That is, frequency features give significance to most naturally occurring signals, e.g., like sound or light. Every periodic signal can be represented as a sum of sine waves of various frequencies, phases, and amplitudes. For example, Eq. 2.1 can be decomposed as the sum of Eq. 2.2 and 2.3, as shown in Figure 2.7.

$$f_1(t) = \sin(t) + 0.2 \cdot \sin(3t) \tag{2.1}$$

$$f_2(t) = \sin(t) \tag{2.2}$$

$$f_3(t) = 0.2 \cdot \sin(3t) \tag{2.3}$$
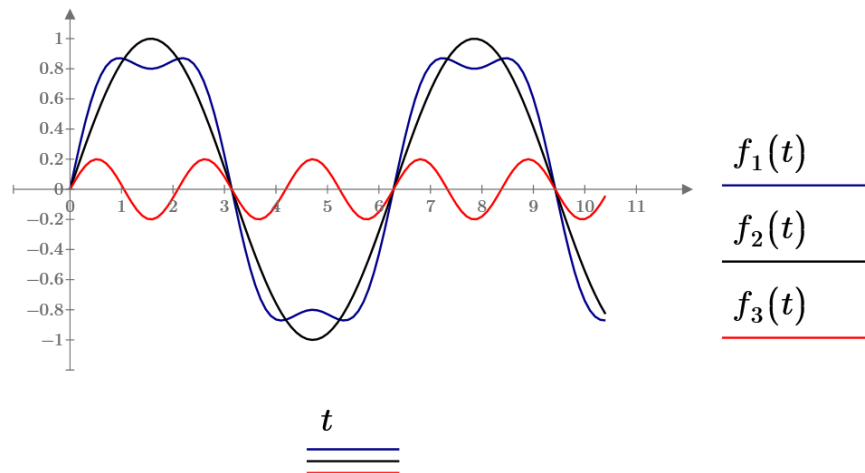
Figure 2.7: Signal decomposition as sum of waves.

The Fourier transform, calculated through Eq. 2.4, allows to extract frequency (ω) magnitude and phase features. The magnitude portion of the Fourier transform for Eq. 2.1 is plotted in Figure 2.8. It can be observed that the magnitude has peaks at one and three, as expected.

$$X(\omega) = \int_{\infty}^{\infty} f(t)e^{-j\omega t}dt \tag{2.4}$$
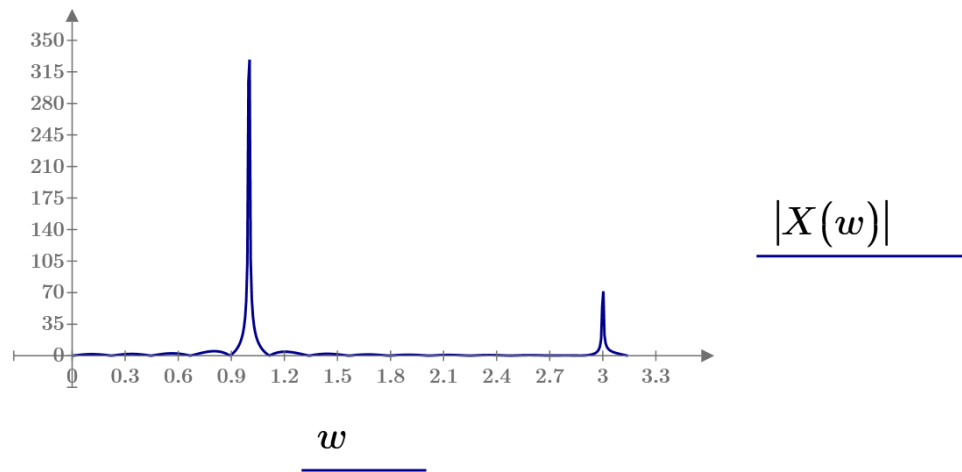
Figure 2.8: Fourier transform magnitude plot (Eq. 2.1).

A limitation to the Fourier transform is that does not illustrate frequency changes over time because it represents data as a sum of sine waves, which extend to infinity. Thus, techniques such as the short-time Fourier transform (STFT) and the wavelet transform have been devised to provide time-frequency signal representations.

2.5.1   Short-Time Fourier Transform

The STFT consists on dividing a long-time signal into shorter portions or bins through a windowing approach and calculating the Fourier transform for each of these bins. This process allows to extract frequency magnitude and phase features and represent them as they change over time. That is, a time-frequency representation. An important condition while performing the STFT is correct window size selection. The window must be long enough to provide enough resolution without compromising the temporal aspects of the signal

2.5.2   Wavelet Transform

A wavelet is a wave-like oscillation with zero mean and finite length. Wavelets come in several form factors, as shown in Figure 2.9, and must be selected depending on the application. In

addition, important wavelet characteristics are scale and shift. Scale refers to how the wavelet is stretched or compressed. A greater scale refers to stretched wavelet, which results in a lower frequency. Shift refers to how a wavelet is delayed or advanced along the signal. A signal's frequency features can be localized in time or space by varying scale and shift parameters.
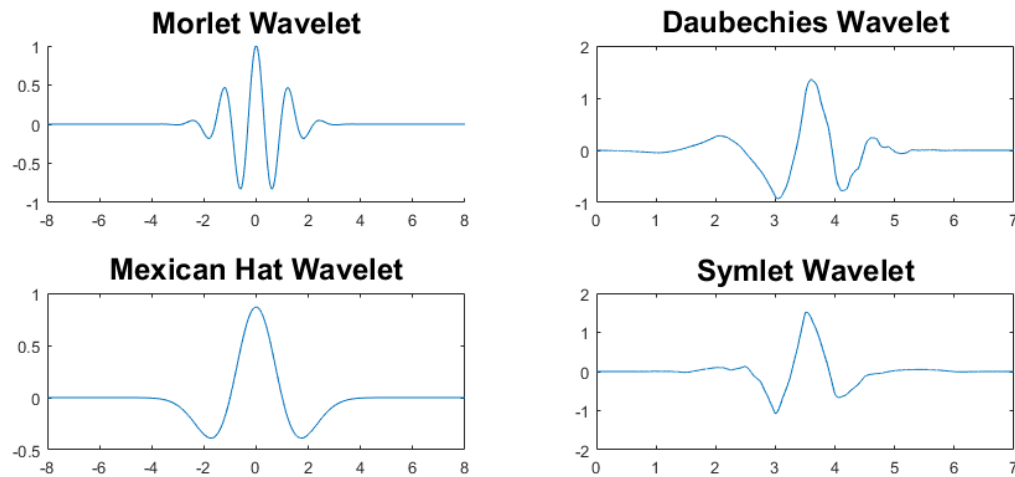
Figure 2.9: Example Morlet wavelet form.

There are two main types of wavelet transforms: the continuous wavelet transform (CWT) and the discrete wavelet transform (DWT). The difference between these two transforms resides on how the shifts and scales are discretized. For a one-dimensional signal (e.g., audio), the output of the CWT are coefficients, which are function of scale (frequency) and shift (time). In MATLAB, the CWT not only allows to scale wavelet frequency by integer powers of two (2^n), or octaves, but also allows scaling within the octaves. The DWT, allows to scale frequencies by a factor of two in multiple levels (Figure 2.10). For each level, the signal is filtered by a high pass and a low pass band. This results on the extraction of frequency coefficients of interest. The process in then repeated for subsequent levels after discarding half the samples per the Nyquist criterion.

Per MathWorks, Inc. (2017-b) the CWT ideal for time-frequency analysis and for filtering of localized frequency components due to its capability for fine-frequency resolution, and the DWT

ideal for signal compression and noise reduction due to its ability to decompose signals in fewer coefficients. Thus, the CWT is used in this study.
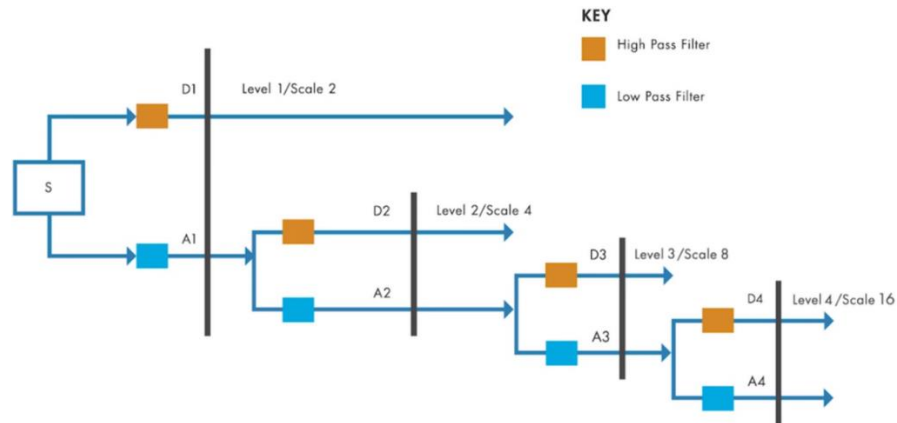
Figure 2.10: Multilevel DWT decomposition (MathWorks, Inc. 2017-b).

2.7   Bayesian Models

An appealing characteristic of Bayesian statistics is the ability to include historical data to perform calculations based on degrees of belief. Bayesian methods have become increasingly relevant in cycle time estimation, productivity estimation, and other situations requiring stochastic modeling. In fact, the Metropolis Algorithm for Monte Carlo has been listed by the IEEE Journal Computing in Science and Engineering as one of the "10 algorithms with the greatest influence on the development and practice of science and engineering in the 20th century" (Dongarra and Sullivan 2000). Using random processes and probabilistic simulations derived from a fraction of the typically-required samples, this algorithm offers an efficient way to seek for answers to problems that may be too complex to solve exactly through a frequentist approach.

2.5.1   Bayesian Models in Manufacturing

There is some prominent research regarding Bayesian models for productivity estimation in the manufacturing industry. Chen, George, and Tardif (2001) proposed a Bayesian approach for

modeling cycle time mean and variance at different levels of work-in-progress. They used Markov Chain Monte Carlo (MCMC) methods, in particular the Gibbs sampling and the Metropolis-Hastings algorithms, to parcel and parametrize cycle time mean versus work-in-progress with linear piecewise functions. Promising results were obtained when comparing the outcomes of this model against a typical non-linear model. Abdoli and Choobineh (2004) led simulations of a resource-sharing, multi-class production environment to compare the performance of Bayes and empirical Bayes methods applied to parametrize different models for flow time forecasting. Their results unequivocally suggest that simpler models reliably yield better forecasts than complex models in which parameters were selected without complete comprehension. More recently, Shen (2008) developed a Bayesian network model for cycle time estimation in the LCD screen defect detection process. Given that defect detection is generally conveyed by human visual inspection, the time required for this process is commonly estimated through complex frequentist statistical models. Nonetheless, Bayesian models, once again, provided a relatively simple and reliable solution.

2.5.2   Bayesian Models in Construction

Because of the complex nature of the construction environment, the industry and academia have long trusted on Bayesian statistical methods in a variety of applications including: modeling workflow for productivity forecasting, analyzing structural resistance to natural forces, and analyzing safety hazards.

Regarding productivity estimation, MCMC-based models have been particularly applicable. Semaan (2016) performed a stochastic productivity analysis of a ready mix concrete batch plant using a queuing model based on Markov chains and a simulation model based on Monte-Carlo-based MicroCyclone modeling software. Results demonstrated that the MicroCyclone simulations

effectively evaluate idleness and yield innovative insight into the impact on plant productivity resulting from changing truck size and quantity. Such findings led to MicroCyclone being used to model numerous activities including: tunneling, paving, bridge construction, bridge redocking, and several other construction operations (Halpin and Riggs 1992, Lutz and Halpin 1992, Pang, Zhang and Hammad 2006).

Concerning structural analysis, precisely Performance Based Design (PBD), Bayesian models are useful to determine the amount of stress that a structure will be subject to when considering natural phenomena. Adeli, et al. (2011) published remarkable insights after performing a probabilistic seismic demand analysis using MCMC methods to simulate the effects on structural performance from parameters with known prior distribution, but no correlation (i.e., earthquakes and economic factors).

Safety in the construction industry deeply relies on providing proper proximity warnings and understanding the workers' responses to such warnings. Looking forward to creating a proactive collision warning system, Zhu, et al. (2016) applied Kalman filtering to predict movement of construction equipment and workers in a construction jobsite. Location estimates from a computer vision framework were provided as input to the filter. Then, the filter generated its own estimates and a corrected location was determined using Kalman gain as a degree of belief. The filters were continuously adjusted based on historical position data and showed incremental effectiveness as more data became available. Luo, et al. (2016) conducted a field experiment to gather location-based data on workers' response rates to varying levels of safety hazard warnings. Bayesian model founded on MCMC methods were applied to get realistic and versatile response rate estimates from simulation because they found that construction jobsites are constantly evolving depending on various factors, like complexity and urgency.

This study aims to profit from the demonstrated versatility of Bayesian models to a field that has been overlooked despite its pressing need and potential for application: cycle time modeling of construction equipment using an audio-based activity identification model as input.

CHAPTER 3

RESEARCH METHODOLOGY

Audio and acceleration data were recorded and processed to evaluate the accuracy of activity classification using support vector machines (SVM). For audio data, frequency feature extraction was executed using the short-time Fourier transform (STFT) and the continuous wavelet transform (CWT). The application of these techniques was compared for activity labeling and for cycle time estimation. For acceleration data, frequency features were only extracted with the CWT and labeling accuracy was compared against audio labeling accuracy. Acceleration data was not used for cycle time estimation. The methodology for audio data processing and acceleration data processing will be discussed separately in sections 3.1 and 3.2, respectively.

3.1   Audio Data Processing

Heavy equipment and tools generate distinctive sound patterns while operating on construction jobsites. It has been proven that audio signals can be processed through machine learning techniques to accurately identify activities performed by such equipment (Cheng, et al., 2017). Taking the output of the previously devised system as direct observation data and using historic data to design Markov-chain-based filter, this study aims at an optimal cycle time forecasting system.

The process for cycle time estimation is summarized in Figure 3.1. To separate the audio signal processing portion from the statistical analysis parts, it will be presented in 2 sections:

- Data Collection and Machine Learning for Classification composed by onsite audio recording and the machine learning framework.

- Cycle Time Forecasting composed by the Markov chain filter and the cycle time estimation algorithm.
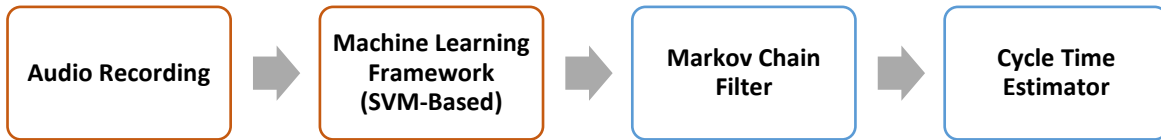
Figure 3.1: Cycle time forecasting framework.

3.1.1 Data Collection and Machine Learning for Classification

The audio signal portion of this study begins with on-site audio recording of construction equipment under normal operation. Then, audio recordings are fed to a machine learning framework consisting of: audio enhancement through a de-noising algorithm, time-frequency representation of the audio signal through the STFT and the CWT for comparison, library generation for posterior labeling using the SVM classifier, and high-level activity label acquisition through a window filtering approach. This process, depicted in Figure 3.2, is detailed in the following sections.

Figure 3.2: Machine learning model for audio signal processing.

3.1.1.1 Audio Recording

Audio data from individual pieces of equipment performing routing actions were taken using the XMOS xCORE-200 multichannel array microphone connected to a laptop computer on site, less than 15 meters away from the sound source of interest. Simultaneously, a video sample was taken 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). The usual setup for audio and video collection is depicted in Figure 3.3.

Figure 3.3: Audio data collection setup.

3.1.1.2   Noise Reduction

Unprocessed audio data signals contained useful audio patterns from heavy equipment mixed with noise from other sound sources found in a jobsite. This noise had to be filtered out to enhance the audio data sets or it could negatively affect the ability of recognizing certain activity patterns. Denoising had to be balanced to effectively remove noise without distorting the signal of interest and it had to adapt considering that noise sources in a jobsite are not constant, e.g., 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 MATLAB implementation. Although this technique was initially devised for speech enhancement, it is versatile enough to be applied in other audio enhancement operations. In principle, the algorithm performs an estimate of the relative level of noise versus signal of interest that adapts quickly in each frame of the signal. Then, a signal smoothing operation in the frequency domain is performed accordingly.

3.1.1.3   Frequency Feature Extraction

Once enhanced, data sets were converted to the time-frequency domain representation using two techniques for comparison: the short-time Fourier transform (STFT) and the continuous wavelet transform (CWT). The STFT consists on dividing a long-time signal into short segments and computing the Fourier transform for each segment; therefore, allowing to extract sinusoidal frequency magnitude and phase content of each segment and represent these features as they change over time. The CWT is a derivation of the Fourier transform that was 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. Unlike sinusoids, wavelets are well localized in time or space. Thus, wavelets are ideal for time-frequency representation of an audio signal.

Wavelets come in different form factors. Selecting size and shape parameters adequately is crucial for each application. An important wavelet characteristic is its scaling factor. A wave's scale factor is inversely proportional to its frequency. That is, scaling a sine wave by 2 (stretching the wave) results in reducing its original frequency by half, or by an octave. For a wavelet, there is a reciprocal relationship between the scale and a constant of proportionality called center frequency. This situation is because a wavelet, unlike a sine wave, has a center frequency and a band-pass characteristic in the frequency domain. The relationship between center frequency and scale is given by Eq. 3.1.

$$F_{eq} = \frac{F_c}{s \cdot T}$$

(3.1)

Where, F_{eq} is equivalent frequency, F_c is the center frequency, s is the scaling factor, and T is the sampling period.

MATLAB was used to implement both techniques. For STFT, a Hanning widow size of 512, a 1024-point discrete Fourier transform, and a 50% overlap (256 samples) were selected. A 512-point window size is optimal because it is long enough to provide a good resolution without compromising the temporal aspects of the signal. For CWT, a bump wavelet was implemented using two sets of parameters: 10 octaves, 24 scales per octave, and a 100-sample shift; and 8 octaves, 32 scales per octave, and a 100-sample shift. A bump wavelet is ideal when it is intended to perform a time-frequency analysis, as is the case with audio signals (MathWorks, Inc. 2017-a).

3.1.1.4   Support Vector Machine

Time-frequency representations of the audio data sets were used to generate a library for posterior activity classification. This library was generated using a support vector machine (SVM) discriminative classifier. The principle for SVM is that, giving it an input of training data for class 1 and class 2 learning, it generates a dividing hyperplane with maximum distance to the training examples (Figure 3.4). Twice this distance is defined as margin. Margin maximization reduces susceptibility to noise while using the SVM generated library to classify new data.
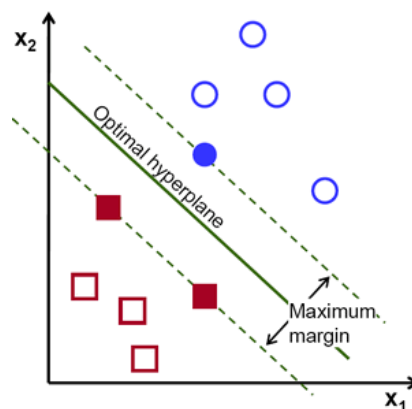
Figure 3.4: Optimal SVM hyperplane (OpenCV 2016).

Let $(\boldsymbol{x_i}, y_i)$ for $i = 1,2, \ldots, n$ be the training data, where $x_i \in \mathbb{R}^d$ is a d-dimensional feature vector, and $y_i \in \{+1, -1\}$ denotes the class of x_i. The equation of an ideal separating hyperplane can be given by Eq. 3.2.

$$\boldsymbol{w} \cdot \boldsymbol{x} + b = 0 \tag{3.2}$$
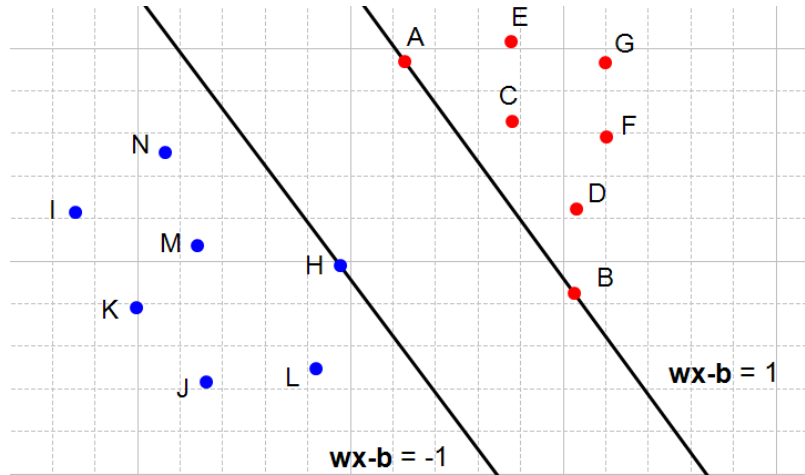
Figure 3.5: Two hyperplanes satisfying the constraints (Kowalczyk 2015).

From Figure 3.5, two constraining hyperplanes, described by Eq. 3.3 and Eq. 3.4, can be selected so that all data points are separated.

$$\boldsymbol{w} \cdot \boldsymbol{x} + b = 1 \tag{3.3}$$

$$\boldsymbol{w} \cdot \boldsymbol{x} + b = -1 \tag{3.4}$$

That means that for each vector $\boldsymbol{x_i}$, either Eq. 3.5 or Eq. 3.6 is fulfilled.

$$\boldsymbol{w} \cdot \boldsymbol{x_i} + b \geq 1 \tag{3.5}$$

$$\boldsymbol{w} \cdot \boldsymbol{x_i} + b \leq -1 \tag{3.6}$$

Multiplying Eq. 3.5 and Eq. 3.6 by y_i, a generalized form for all i can be obtained, as seen in Eq. 3.7.

$$y_i(\boldsymbol{w} \cdot \boldsymbol{x_i} + b) \geq 1 \tag{3.7}$$

It can be proven that the maximum margin is given by Eq. 3.8.

$$m = \frac{2}{\|\boldsymbol{w}\|} \tag{3.8}$$

Thus, maximizing the margin involves minimizing $\|\boldsymbol{w}\|$ subject to Eq. 3.6.

Of course, actual training data is usually multidimensional and hyperplane determination requires a complex optimization process. To make the binary class separation easier, it is first necessary to perform mapping operation $\Phi : \mathbb{R}^d \to \mathcal{H}$, where $\mathcal{H}$ is a high dimensional Hilbert space. Increasing the dimensionality of data improves its resemblance to a linearly separable data set. For this study, mapping is performed through the radian basis function (RBF) kernel, depicted in Eq. 3.9.

$$K(\boldsymbol{x}, \boldsymbol{x'}) = \exp(-\gamma \|\boldsymbol{x} - \boldsymbol{x'}\|^2) \tag{3.9}$$

The SVM training operation was performed using the MATLAB Statistics and Machine Learning Toolbox. Through the *fitcsvm* command, a variant to the optimization problem is resolved:

$$\begin{array}{c} min \\ \boldsymbol{w}, b, \xi \end{array} \quad \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{n} \xi_i \tag{3.10}$$

subject to,

$$y_i(\boldsymbol{w} \cdot \boldsymbol{x_i} + b) \geq 1 - \xi_i \qquad \text{for } i = 1, 2, \dots, n \tag{3.11}$$

$$\xi_i \geq 0 \qquad \text{for } i = 1, 2, \dots, n \tag{3.12}$$

Where, ξ_i is a set of slack variables introduced to relax the constraints in case that the classes can't be separated and C is an overfitting prevention parameter.

Four audio segments of the construction equipment performing a major activity were 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 two to six seconds long and included only the STFT frequency or wavelet coefficient magnitude portion. To guarantee correct SVM parameter selection, ten-fold cross validation was used. All γ, ξ_i, and C were set to be automatically optimized by the training algorithm.

3.1.1.5   Window Filtering

Once an SVM library had been generated for a specific construction equipment, it was 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 segment to the next. Therefore, a window filtering algorithm was implemented to smooth out the classified output. The window filtering parameters are small window size, large window size, and 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. Window sizes varied, but usual size for the small window was one-quarter of a second and for large window was one to three seconds.

3.1.1.6   Confusion Matrix

Once labels were generated, the results were evaluated using a confusion matrix by comparing the predicted labels vs. the manually-classified correct labels. Refer to the example confusion

matrix portrayed in Table 3.1, cell TP1 is used to indicate true positive classification for activity 1, cell FP2 is used to indicate false positive classification for activity 2, cell FP1 is used to indicate false positive classification for activity 1, and cell FP2 is used to indicate false positive classification for activity 2.

Table 3.1: Example confusion matrix.

Example		Correct Label	
		Act 1	Act 2
Predicted Label	Act 1	Cell TP1	Cell FP1
	Act 2	Cell FP2	Cell TP2

3.1.2 Cycle Time Forecasting

Regardless of the nature of the project, it is likely to involve earthmoving and material moving cyclic operations. As shown in Table 3.2, these actions are performed through a sequence of major and minor activities. Major activities are value-generating activities and minor activities are necessary transition activities. Using the output from the previous section, cycle time estimation is executed after the Markov chain filter. More details are provided in the following sections.

Table 3.2: Typical actions performed by heavy equipment.

Equipment	Action	Typical Activity Sequence	Type
Excavator/ Loader/ Dozer (dozer less effective)	Excavating/ Moving material/ Backfilling/ Truck loading	Digging	Major
		Swinging or maneuvering	Minor
		Dumping	Major
		Swinging or maneuvering	Minor
Excavator/ Loader	Compacting/ Demolishing	Compressing with bucket	Major
		Swinging or maneuvering	Minor
Grader/ Dozer/ Loader (loader less effective)	Grading/ Ripping/ Clearing/ Blending	Pushing material with blade/bucket	Major
		Reversing or maneuvering	Minor

3.1.2.1 Markov Chain Filter

The output from the audio framework was not sufficiently smooth to estimate cycle times accurately. Therefore, Markov chains were incorporated to include ground truth statistical data

into activity labeling. To design a suitable Markov chain, concepts like time-frequency bins per second, decisions while in act, and calls for act had to be devised.

The number of frequency bins per second for the STFT is obtained through Eq. 3.13. If the sampling frequency is 44100 Hz, the window size is 512 samples, and 256 samples are overlapped, the number of time-frequency bins per second is 172.26. If the CWT were used, the number of bins per second would simply be the original sampling frequency divided by the shifting parameter, or 441 time-frequency bins per second.

$$BPS = \frac{Sampling\ frequency}{Window\ size - Overlapped\ samples} \tag{3.13}$$

The SVM classifier labels each bin so the time elapsed in each activity multiplied by the number time-frequency bins represents decisions taken while in each activity. From manually-labeled activities, the total time spent while performing major activities (Act 1) and minor activities (Act 2) was calculated and then multiplied by the number of time-frequency bins in one second. The number of calls for each activity is the number of transitions from the previous activity. The probability of the state changing to Act 2 given that it is Act 1 is equivalent to the calls for Act 2 divided by the number of decisions taken while in Act 1. The probability of the state being Act 1 and keep being Act 1 is the complement. This is illustrated by Eq. 3.14 to 3.17.

$$P(Act\ 2\ |\ Act\ 1) = \frac{Calls\ for\ Act\ 2}{Decisions\ for\ Act\ 1} \tag{3.14}$$

$$P(Act\ 1\ |\ Act\ 1) = 1 - P(Act\ 2\ |\ Act\ 1) \tag{3.15}$$

$$P(Act\ 1\ |\ Act\ 2) = \frac{Calls\ for\ Act\ 1}{Decisions\ for\ Act\ 2} \tag{3.16}$$

$$P(Act\ 2\ |\ Act\ 2) = 1 - P(Act\ 1\ |\ Act\ 2) \tag{3.17}$$

Table 3.3: Ground truth data for JD 700J assuming STFT for frequency feature extraction.

Activity	Start (seconds)	Elapsed time	Call Act 1	Call Act 2	Act 1 time	Act 2 time	Decisions 1	Decisions 2
Pushing soil with blade	0	35	NA		35		6029	
Reversing	35	13		1		13		2239
Pushing soil with blade	48	25	1		25		4306	
Reversing	73	13		1		13		2239
Pushing soil with blade	86	44	1		44		7579	
Reversing	130	28		1		28		4823
Pushing soil with blade	158	23	1		23		3962	
Reversing	181	16		1		16		2756
Pushing soil with blade	197	46	1		46		7924	
Reversing	243	23		1		23		3962
Pushing soil with blade	266	31	1		31		5340	
Reversing	297	16		1		16		2756
Pushing soil with blade	313	27	1		27		4651	
Reversing	340	14		1		14		2412
Pushing soil with blade	354	5	1		5		861	
End	359							
TOTAL			7	7	236	123	40653	21188

A typical arrangement of ground truth data for Markov chain design using the STFT approach is depicted in Table 3.3. The total time that the construction equipment spent on performing major activities and minor activities was manually labeled. This is indicated in the columns Act 1 Time and Act 2 Time. Multiplying these by the number of bins per second (i.e., 172.26) produces the values in columns Decisions 1 and Decisions 2. The number of calls for each activity is simply the number of transitions from Act 1 to Act 2, and vice versa. Using this data, the elements of the Markov chain are:

While in Act 1,

$$P(Act\ 2\ |\ Act\ 1) = \frac{Calls\ for\ Act\ 2}{Decisions\ for\ Act\ 1} = \frac{7}{40653} = 0.00017219 \rightarrow 0.017\%$$

$$P(Act\ 1\ |\ Act\ 1) = 1 - P(Act\ 2\ |\ Act\ 1) = 0.99982781 \rightarrow 99.983\%$$

While in Act 2,

$$P(Act\ 1\ |\ Act\ 2) = \frac{Calls\ for\ Act\ 1}{Decisions\ for\ Act\ 2} = \frac{7}{21188} = 0.000330 \rightarrow 0.033\%$$

$$P(Act\ 2\ |\ Act\ 2) = 1 - P(Act\ 1\ |\ Act\ 2) = 0.999670 \rightarrow 99.967\%$$

The Markov chain matrix for the JD 700J dozer using the state-dependent probability
distributions is depicted in Table 3.4. A graphical representation of the Markov process is depicted in

Figure 3.6.

Table 3.4: Markov matrix for JD 700J.

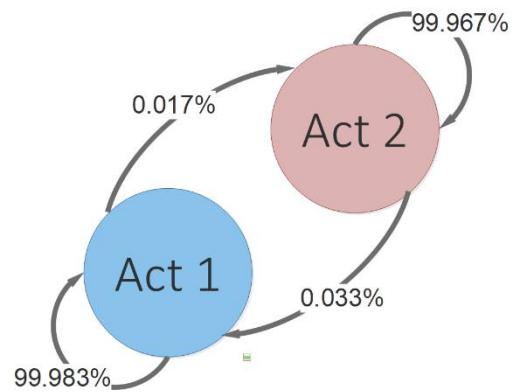| John Deere 700J		Next State	
		Act1	**Act2**
Current State	**Act1**	0.999828	0.000172
	Act2	0.000330	0.999670

Figure 3.6: Two-state Markov process for JD 700J.

The flow diagram for the Bayesian filter is shown in Figure 3.7. The audio portion (sensor data) is depicted in blue, the Markov process portion is depicted in red, the current state is depicted in green, and decision boxes are depicted in grey. The predicted state for the Markov chain is the one with highest probability in the Markov process. The accuracy for the prediction is the probability by which it was predicted. Likewise, the accuracy for the SVM-predicted state is taken as the percentage over the widow filter by which it was determined. The next state is the one with greater accuracy, either the SVM-predicted state or the Markov process state. The exponent (n) of the Markov process is reset if the current state diverges from the previous state. Otherwise, the Markov matrix is elevated to the next power (n+1) for the following step.
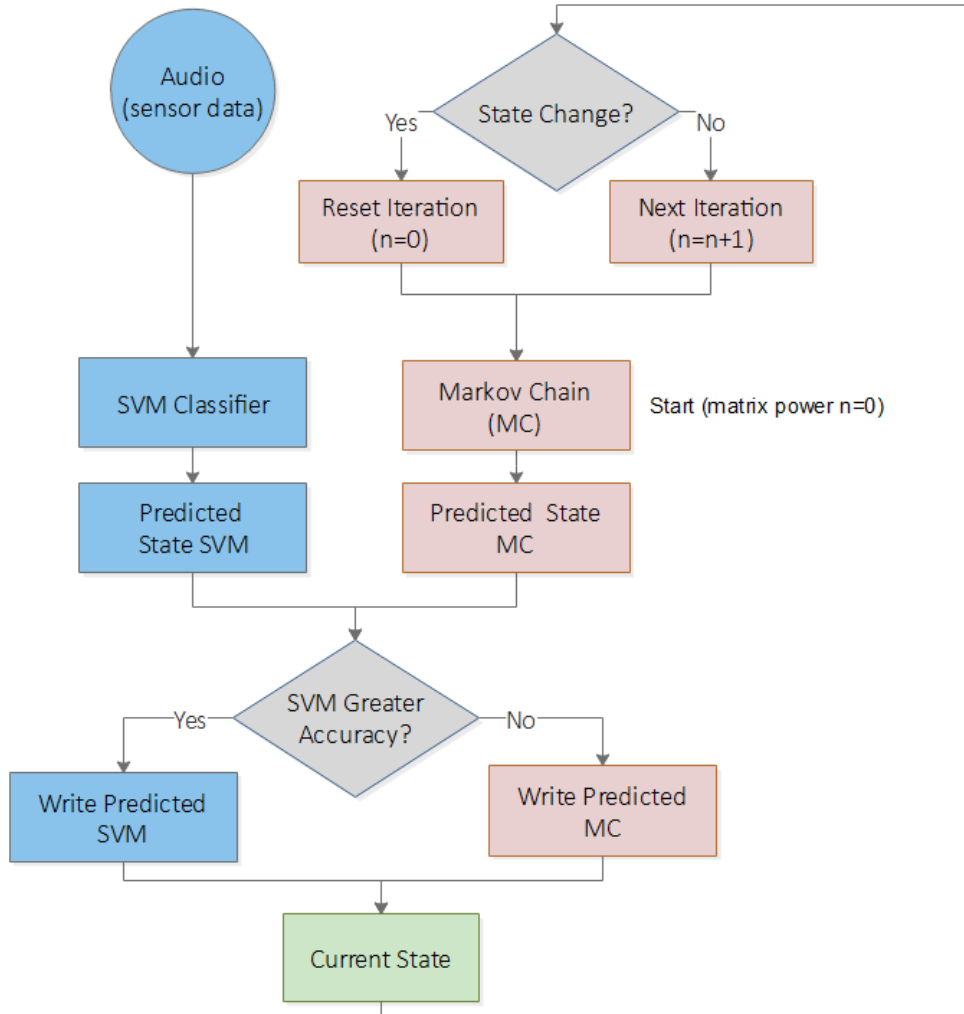
Figure 3.7: Adaptive Markov filter process diagram.

3.1.2.2 Cycle Time Estimator

If cycle time for a specific action can be accurately measured, then it can be used along with manufacturer data to determine the equipment productivity. A machine work cycle is a succession of major and minor activities, as shown in Table 3.2. Cycle time is the time elapsed during such succession. Therefore, the cycle time estimator was designed in MATLAB to scan through the labeled audio signal, count continuous activities by type, and determine the average time elapsed on each cycle.

3.2   Accelerometer Data Processing

Mobile phone micro-electro-mechanical-systems (MEMS) accelerometers were used to collect data and process it through a model similar to that used for audio signals, as depicted in Figure 3.8. The objective was to evaluate activity labeling accuracy and compare to that achieved with audio signals using the confusion matrix approach. The details about the MEMS model will be discussed in the following paragraphs.

Figure 3.8: Machine learning model for MEMS accelerometer data processing.

Data was collected using a mobile phone interfaced to a laptop computer on site via Wi-Fi using the MATLAB Support Package for Android Devices. The mobile device was fixed on board heavy equipment with a support arm, as depicted in Figure 3.9. Video and audio data was jointly collected for reference and comparison.

Figure 3.9: MEMS data collection setup.

Accelerometer data obtained from cell phone devices was captured for three axes, as shown in Figure 3.10. To standardize this data and disregard the orientation of the device, it was preferred

to work with the magnitude of the acceleration data and remove any constant effects (e.g., gravity) by subtracting the mean from the data. An example standardized data set is depicted in Figure 3.11.
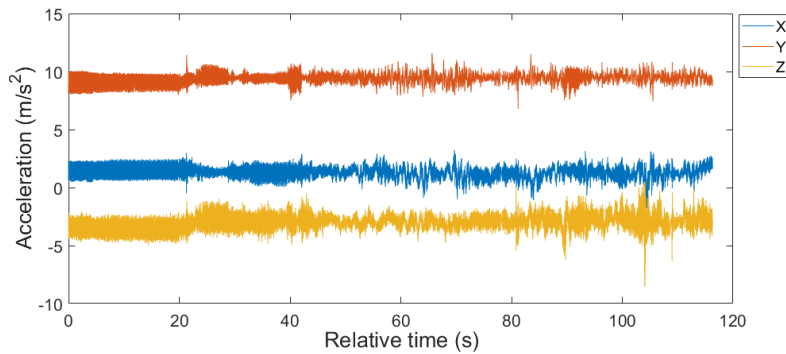

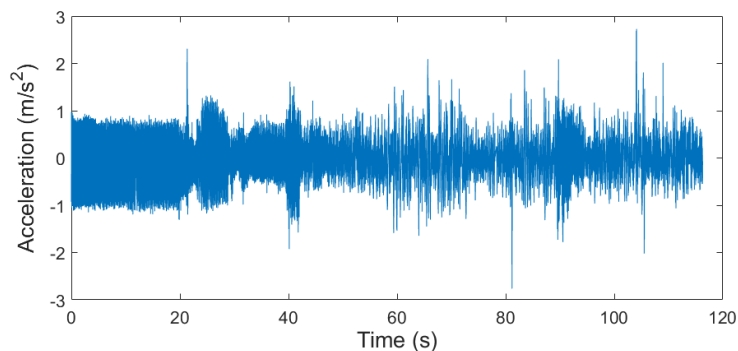Figure 3.10: Three axis representation of acceleration data.

Figure 3.11: Standardized acceleration data.

Standardized accelerometer data contained information of interest along with noise. However, noise was not as directly influenced by external sources because the sensors were mounted directly on board the equipment. It was observed that a third-order Butterworth low-pass filter with a cutoff frequency of half the sampling frequency sufficed to eliminate noise and aliasing. Given that a maximum sampling frequency of 100 Hz was achievable with these MEMS devices, a 50 Hz cutoff frequency was selected.

For frequency feature extraction, the CWT was employed using a bump wavelet with 4 octaves, and 48 scales per octave. CWT was preferred over STFT because it allowed a higher

resolution time-frequency representation than the windowing approach required for STFT. That is, with the CWT an array of 100 columns can be obtained for one second of data (one column per sample), as opposed the array size achievable using a few samples per window.

For SVM training, four accelerometer data segments of the construction equipment performing a major activity were 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 two to six seconds long and included the CWT scale coefficient magnitude along with its corresponding label. To guarantee correct SVM parameter selection, ten-fold cross validation was used. All γ, ξ_i, and C were set to be automatically optimized by the training algorithm.

Once an SVM library had been generated for a specific construction equipment, it was 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 segment to the next. Therefore, a window filtering algorithm was implemented to smooth out the classified output. The window filtering parameters are small window size, large window size, and 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. Window sizes varied, but usual size for the small window was one-quarter of a second and for large window was one to three seconds.

CHAPTER 4

RESULTS AND DISCUSSION

Results are presented and discussed in five sections: first, the support vector machine (SVM) labeling results obtained from processing audio signals through the continuous wavelet transform (CWT with 10 octaves and 24 scales per octave) versus the short-time Fourier transform (STFT) are compared; second, the SVM labeling results obtained from processing audio signals through the CWT (8 octaves and 32 scales per octave) versus the STFT are compared; third, the SVM labeling results obtained from processing audio data are compared against the results obtained from processing active sensor data; fourth, the Markov filter is applied to audio SVM labels after CWT and STFT feature extraction to evaluate single-day cycle time estimation accuracy; and, finally, the Markov filter is applied to audio SVM labels after STFT feature extraction to evaluate multiple-day cycle time estimation accuracy.

4.1   Audio - CWT (NO: 10 and SO: 24) vs. STFT

The performance of the audio signal processing framework for the first CWT configuration (10 octaves and 24 scales per octave) versus the STFT (1024 frequency points, 512-sample window, and 256 overlapped samples) was evaluated using recordings taken at local jobsites for the following equipment: 1) John Deere 700J dozer, 2) John Deere 670G grader, 3) JCB 3CX backhoe excavator, and 4) Komatsu PC200 excavator. Each recording was submitted to denoising, time-frequency feature extraction via the STFT and the CWT, and SVM training using 10 to 30 seconds of data for activity 1 (major activity) and 10 to 30 seconds of data for activity 2 (minor activity). Then, activity classification and filtering were performed to an independent audio segment. The classified labels are shown graphically in Figures 4.1, 4.3, 4.5, and 4.7. In each

figure, the upper plot shows the labels obtained using the features extracted using the CWT, the middle plot shows the labels obtained using the features extracted using the STFT, and the bottom plot shows the observed, correct labels. Major activities are represented by a high position and minor activities are represented by a low position. A 30-second portion of the CWT and the STFT of each audio file is depicted in Figures 4.2, 4.4, 4.6, and 4.8. A black line depicting major and minor activities was plotted over the figures to provide physical significance to the frequency magnitude graphs. The visual time-frequency representations had to be cropped due the computational cost of plotting the CWT. That is, the CWT for a 300-second audio file, with a sampling frequency of 441000 Hz, and 240 scales yields an array with over 3 billion cells, as opposed to roughly 52 thousand cells required for the STFT with the current configuration.

The comparison graphs for the John Deere 700J dozer after implementing the CWT with 10 octaves and 24 scales per octave versus the STFT are shown in Figures 3.1 and 3.2. Major activities included pushing soil and duping while minor activities included maneuvering and reversing.
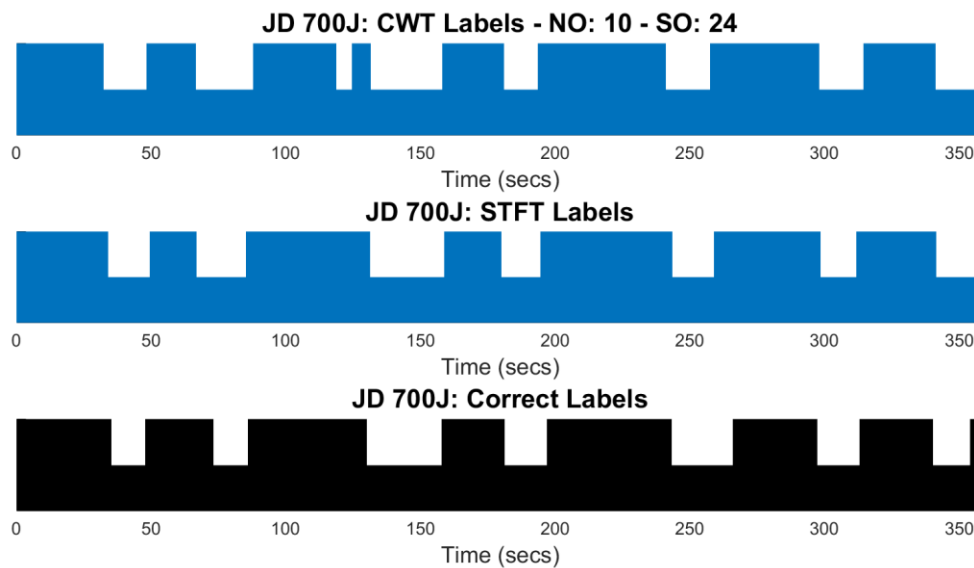

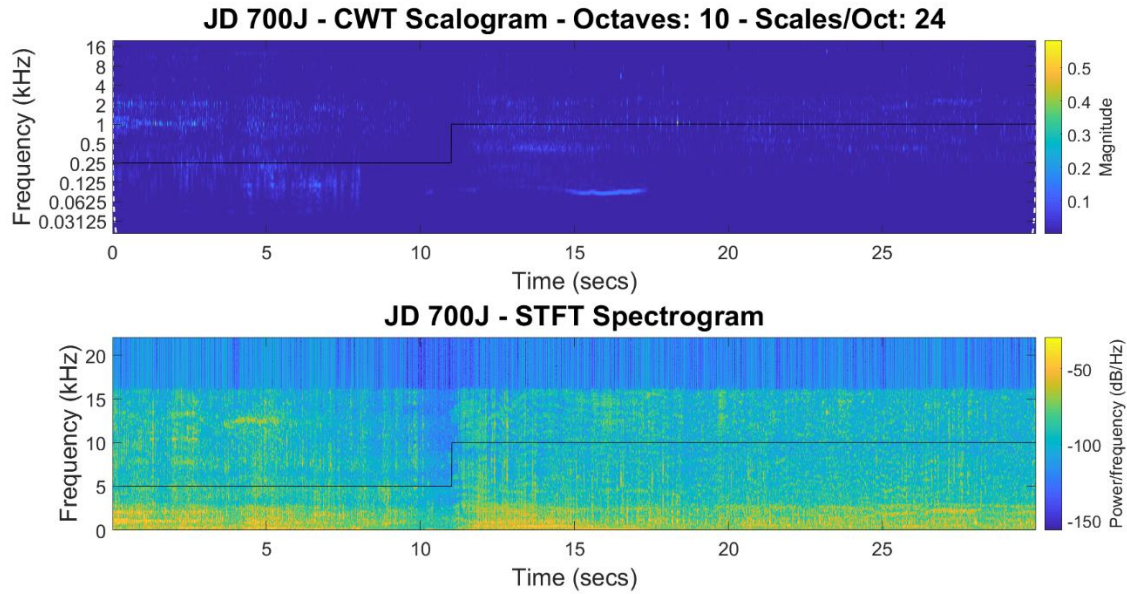Figure 4.1: JD 700J – CWT 10/24 vs. STFT labeling comparison.

Figure 4.2: JD 700J– CWT 10/24 scalogram vs. STFT spectrogram comprison.

The comparison graphs for the John Deere 670G grader after implementing the CWT with 10 octaves and 24 scales per octave versus the STFT are shown in Figures 4.3 and 4.4. Major activities included grading and clearing surface soils while minor activities included maneuvering and reversing.
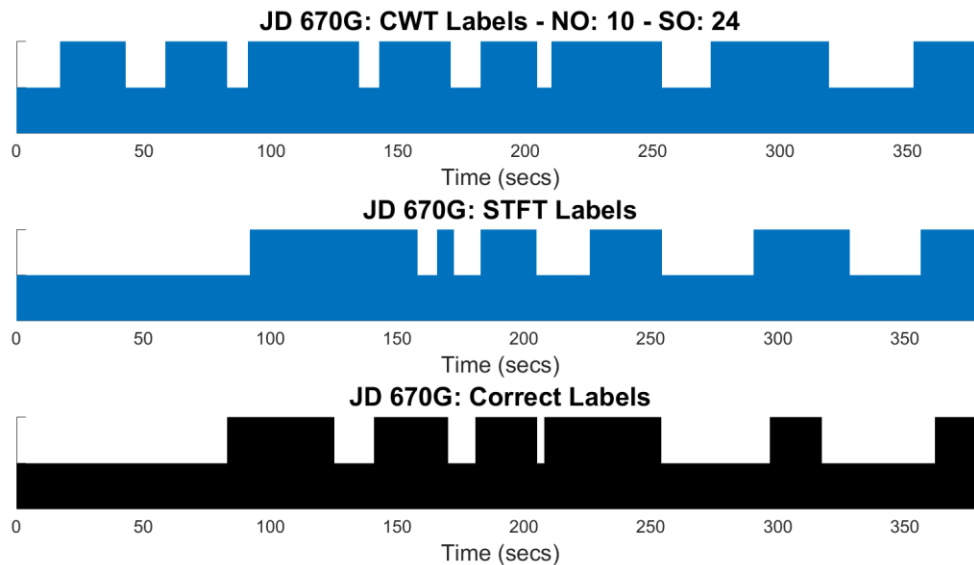

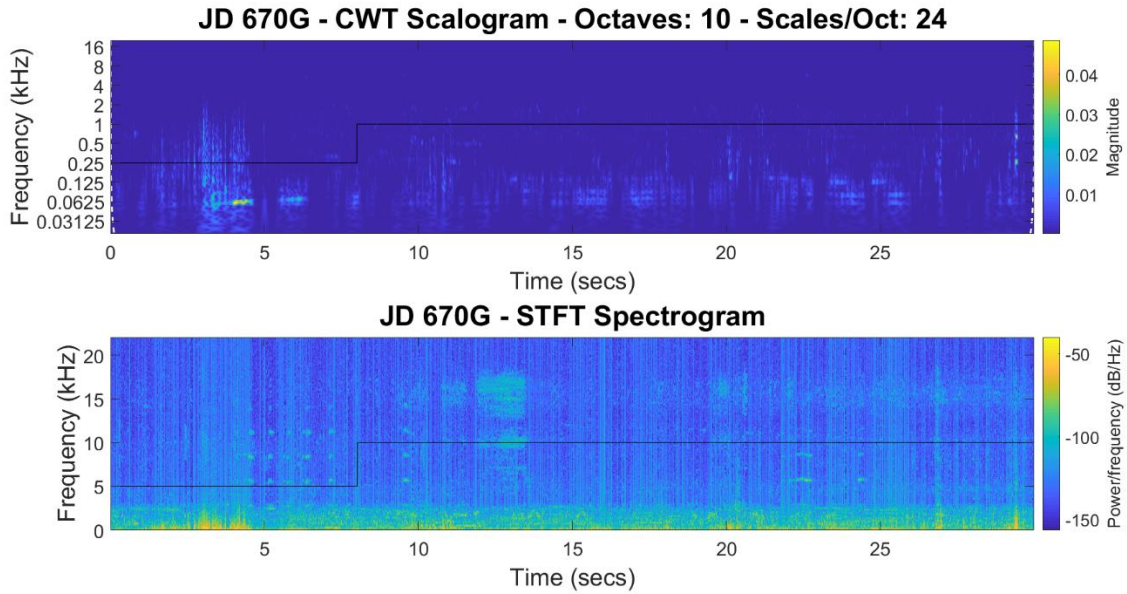Figure 4.3: JD 670G – CWT 10/24 vs. STFT labeling comparison.

Figure 4.4: JD 670G– CWT 10/24 scalogram vs. STFT spectrogram comprison.

The comparison graphs for the JCB 3CX backhoe excavator after implementing the CWT with 10 octaves and 24 scales per octave versus the STFT are shown in Figures 4.5 and 4.6. Major activities included excavating, scooping, and dumping while minor activities included maneuvering, swinging, and idle times.
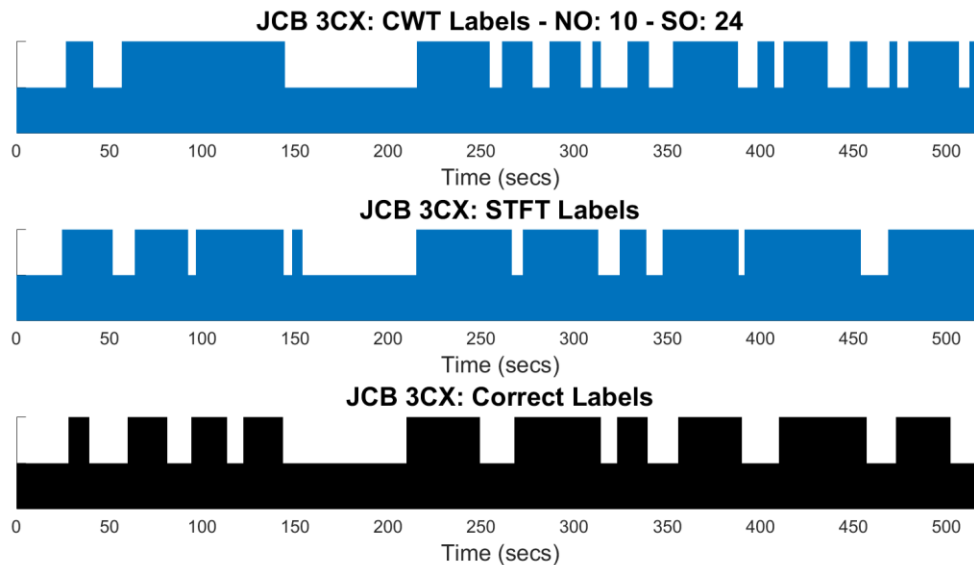

Figure 4.5: JCB 3CX – CWT 10/24 vs. STFT labeling comparison.

Figure 4.6: JCB 3CX– CWT 10/24 scalogram vs. STFT spectrogram comprison.

The comparison graphs for the Komatsu PC200 excavator after implementing the CWT with 10 octaves and 24 scales per octave versus the STFT are shown in Figures 4.7 and 4.8. Major activities included excavating, scooping, and dumping while minor activities included maneuvering, swinging, and idle times.

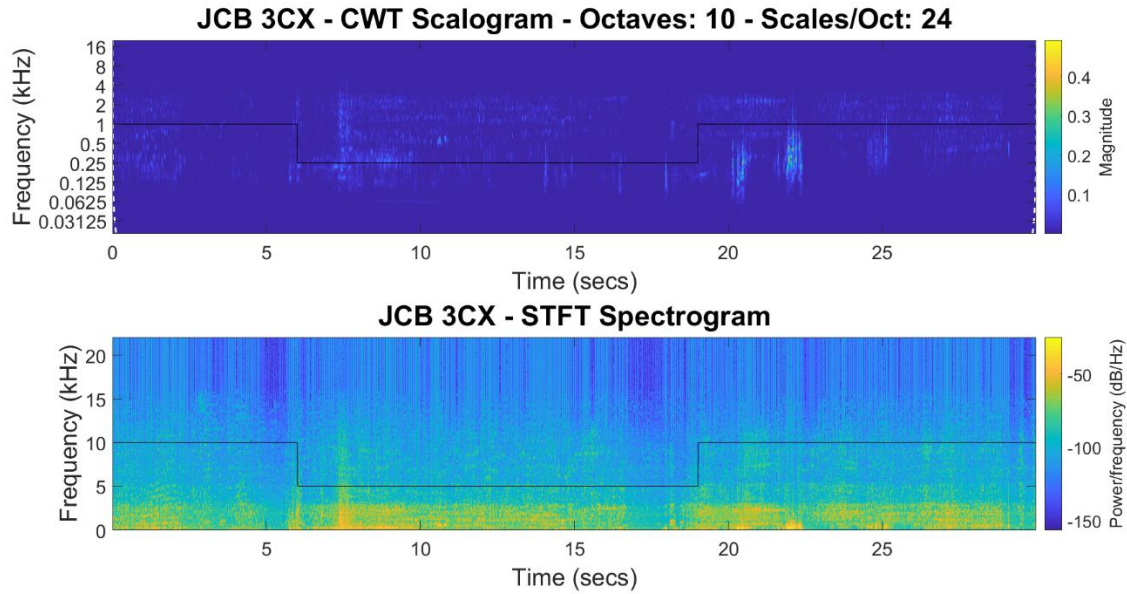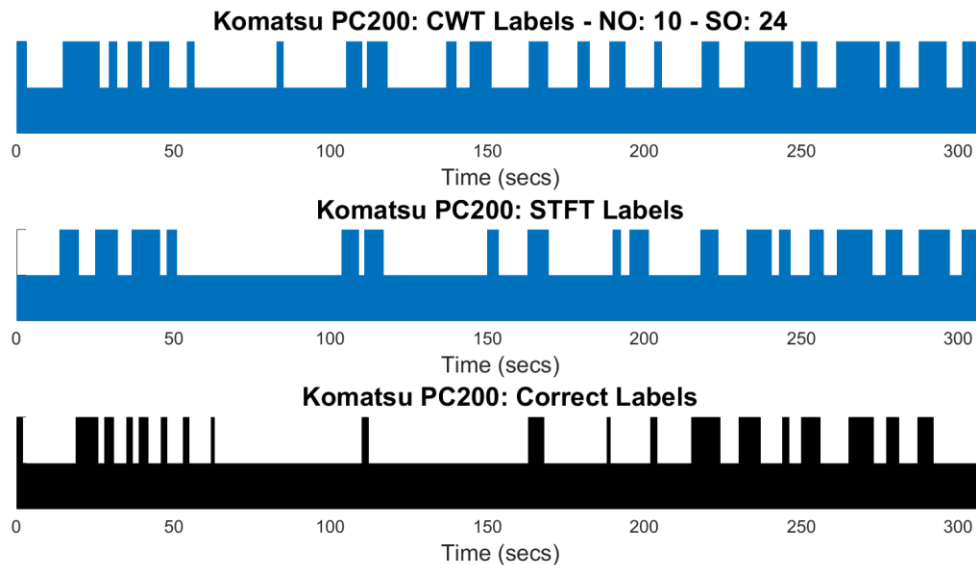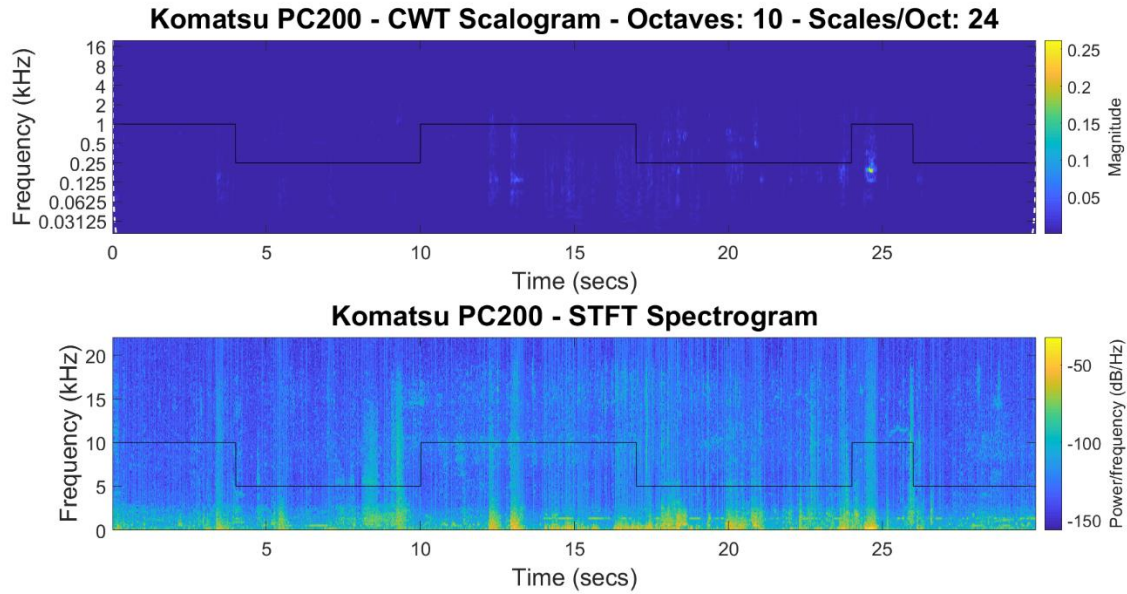Figure 4.7: Komatsu PC200 – CWT 10/24 vs. STFT labeling comparison.

Figure 4.8: Komatsu PC200– CWT 10/24 scalogram vs. STFT spectrogram comprison.

To provide a more objective comparison for labeling accuracies, the confusion matrix true

positives for major activities (Act 1) and minor activities (Act 2) are summarized side by side in

Table 4.1. Refer to Table 3.1 for an explanation about confusion matrices and true positives.

Consider that a balance in true positives is better than having a high accuracy in one class at the

cost of the accuracy of the other class. By careful observation, it can be determined that the STFT

provides slightly better results than the CWT

Table 4.1: CWT 10/24 vs. STFT true positive classification accuracy comparison.

	CWT 10/24		STFT	
	Act 1	Act 2	Act 1	Act 2
JD 700J	87.30%	88.37%	93.44%	87.53%
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%

Regarding CWT configuration, per MathWorks, Inc. (2017-a), using a smaller number of

octaves with a higher resolution within the octaves is preferable when the features of interest

primarily are contained in higher frequencies. The 10-octave CWT scalograms presented in this

section include frequencies less than 125 Hz, which can be observed to have negligible magnitudes.

4.1   Audio - CWT (NO: 8 and SO: 32) vs. STFT

The performance of the audio signal processing framework for the second CWT configuration (8 octaves and 32 scales per octave) versus the STFT was evaluated using recordings taken at local jobsites for the following equipment: 1) John Deere 700J dozer, 2) John Deere 670G grader, 3) JCB 3CX backhoe excavator, 4) Komatsu PC200 excavator, and 5) Komatsu 39PX dozer. SVM training and classification was performed as in the previous section. The classified labels are shown graphically in Figures 4.9, 4.11, 4.13, 4.15, and 4.17. A 30-second portion of the CWT and the STFT of each audio file is depicted in Figures 4.10, 4.12, 4.14, 4.16, and 4.18.

The comparison graphs for the John Deere 700J dozer after implementing the CWT with 8 octaves and 32 scales per octave versus the STFT are shown in Figures 4.9 and 4.10. Major activities included pushing soil while minor activities included maneuvering and reversing.
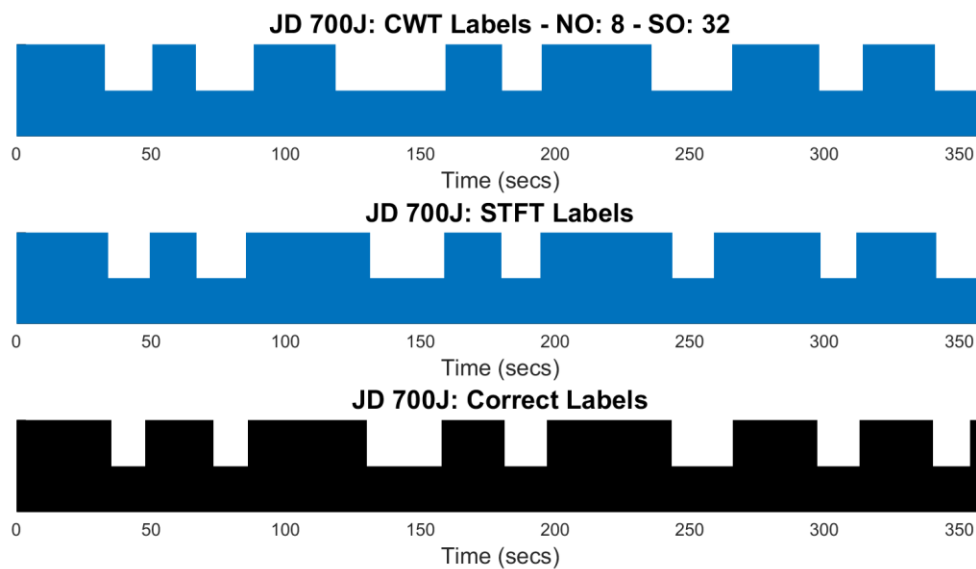

Figure 4.9: JD 700J – CWT 8/32 vs. STFT labeling comparison.

Figure 4.10: JD 700J– CWT 8/32 scalogram vs. STFT spectrogram comprison.

The comparison graphs for the John Deere 670G grader after implementing the CWT with 8 octaves and 32 scales per octave versus the STFT are shown in Figures 4.11 and 4.12. Major activities included grading and clearing surface soils while minor activities included maneuvering and reversing.

Figure 4.11: JD 670G – CWT 8/32 vs. STFT labeling comparison.
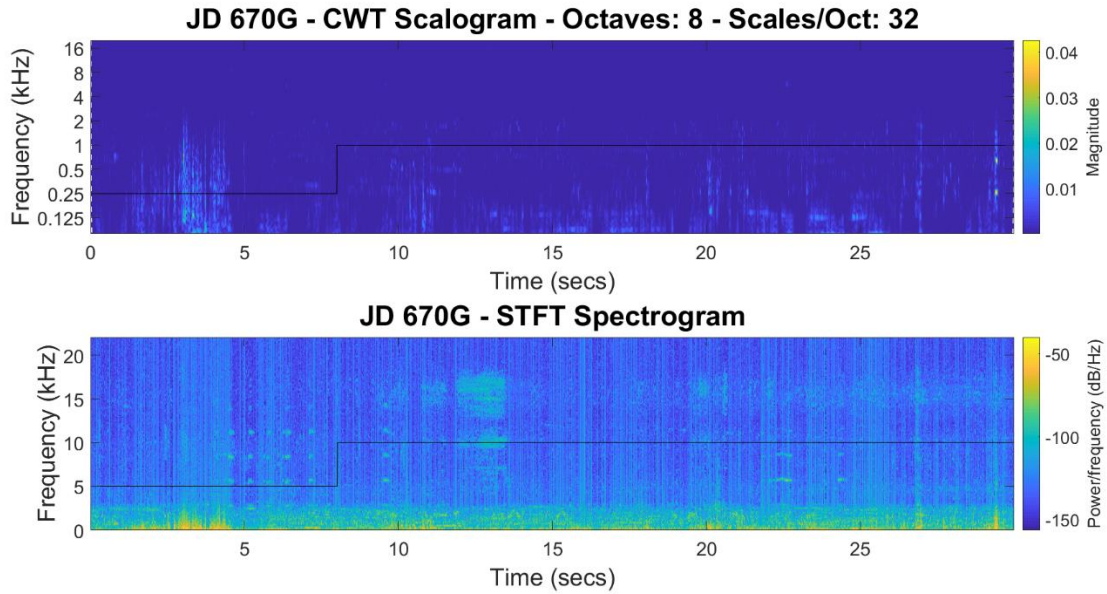
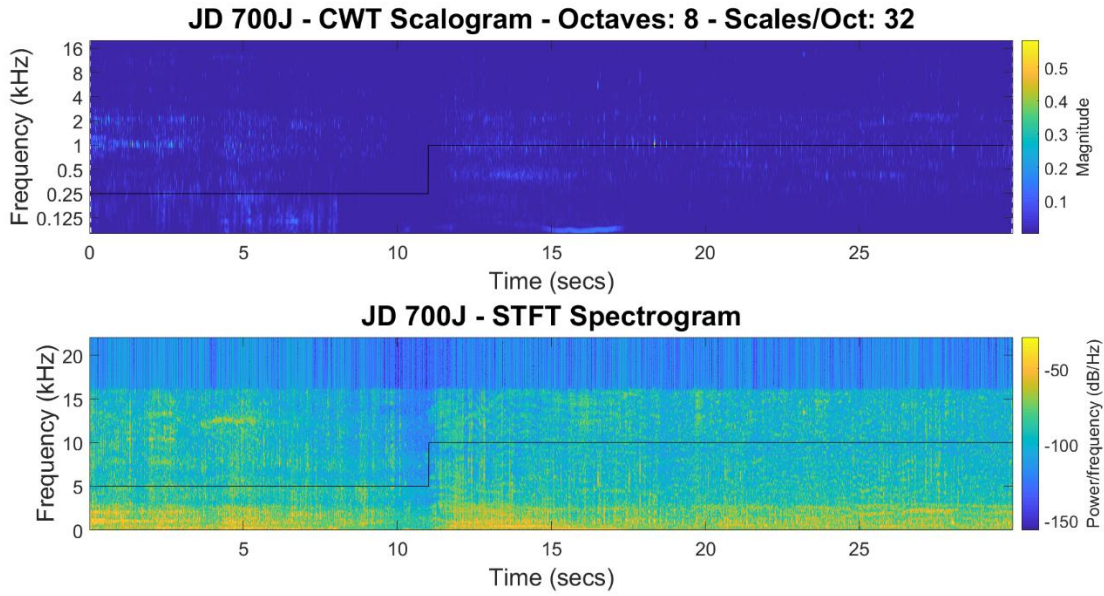Figure 4.12: JD 670G– CWT 8/32 scalogram vs. STFT spectrogram comprison.

The comparison graphs for the JCB 3CX backhoe excavator after implementing the CWT with 8 octaves and 32 scales per octave versus the STFT are shown in Figures 3.13 and 3.14. Major activities included excavating, scooping, and dumping while minor activities included maneuvering, swinging, and idle times.

Figure 4.13: JCB 3CX – Labeling CWT 8/32 vs. STFT labeling comparison.

Figure 4.14: JCB 3CX– CWT 8/32 scalogram vs. STFT spectrogram comprison.

The comparison graphs for the Komatsu PC200 excavator after implementing the CWT with 8 octaves and 32 scales per octave versus the STFT are shown in Figures 4.15 and 4.16. Major activities included excavating, scooping, and dumping while minor activities included maneuvering, swinging, and idle times.

Figure 4.15: Komatsu PC200– Labeling CWT 8/32 vs. STFT labeling comparison.

Figure 4.16: Komatsu PC200– CWT 8/32 scalogram vs. STFT spectrogram comprison.

The comparison graphs for the Komatsu 39PX dozer after implementing the CWT with 8 octaves and 32 scales per octave versus the STFT are shown in Figures 4.17 and 4.18. Major activities included pushing soil and duping while minor activities included maneuvering and reversing.

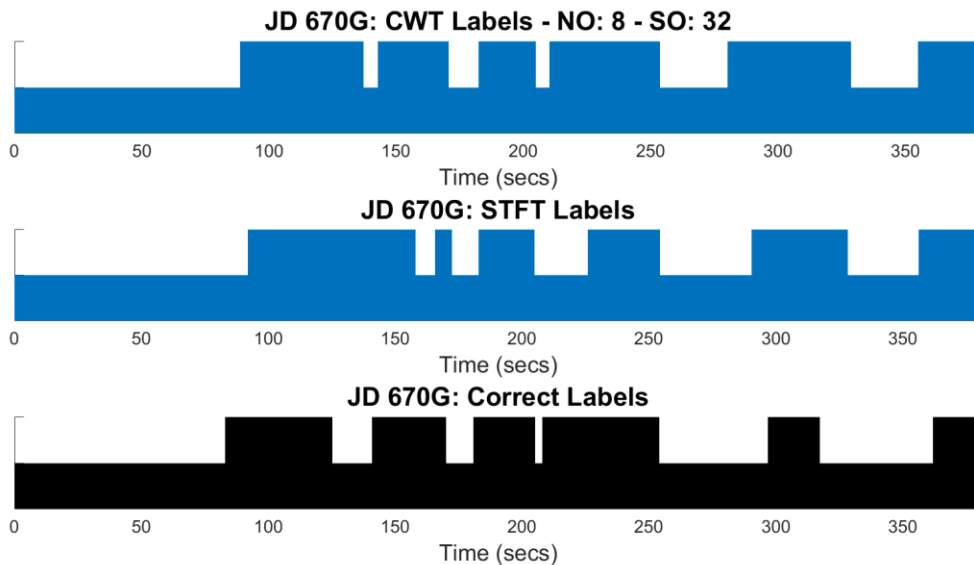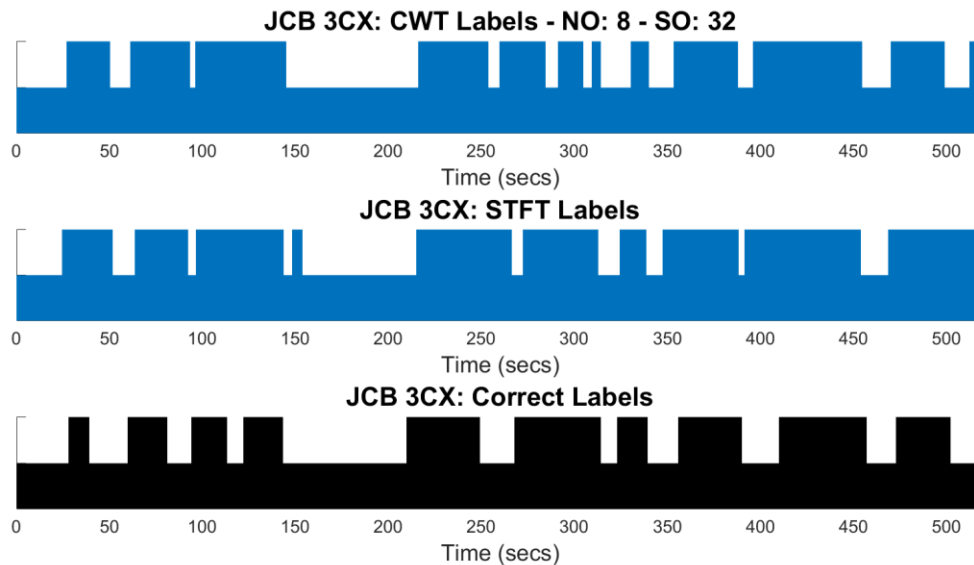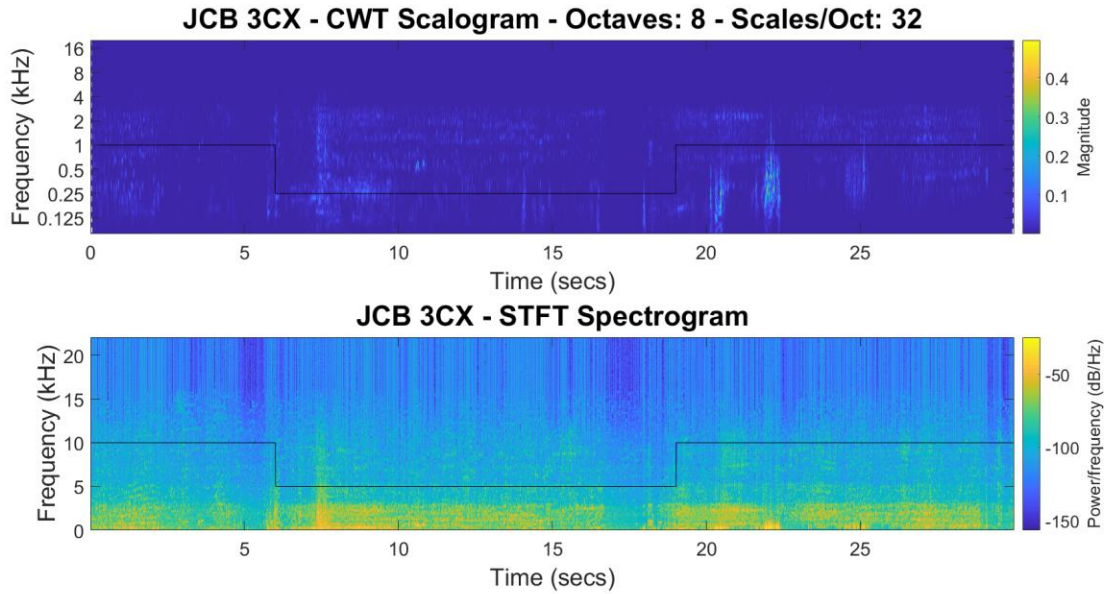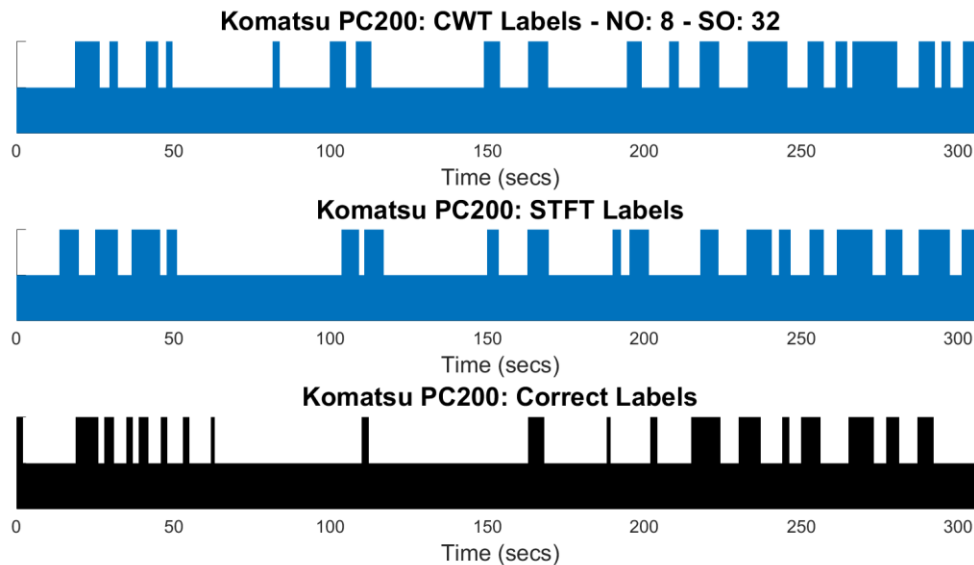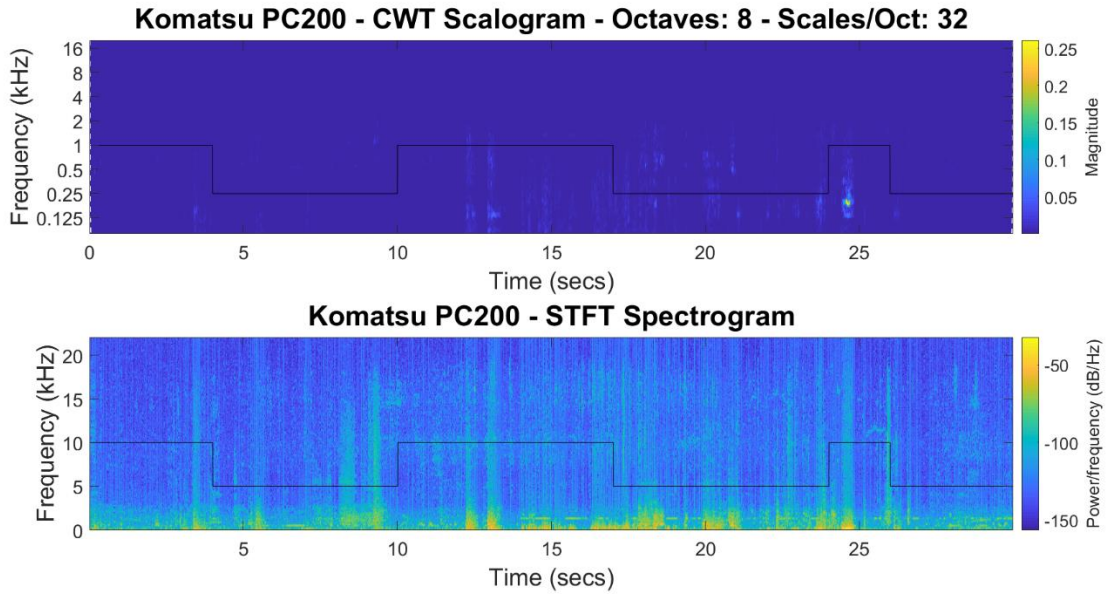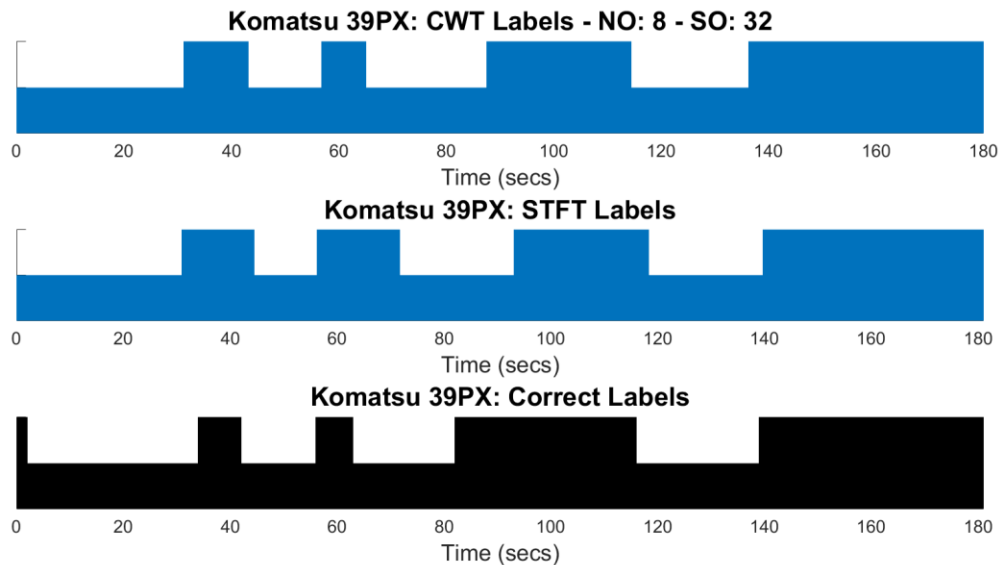Figure 4.17: Komatsu 39PX – Labeling CWT 8/32 vs. STFT labeling comparison.
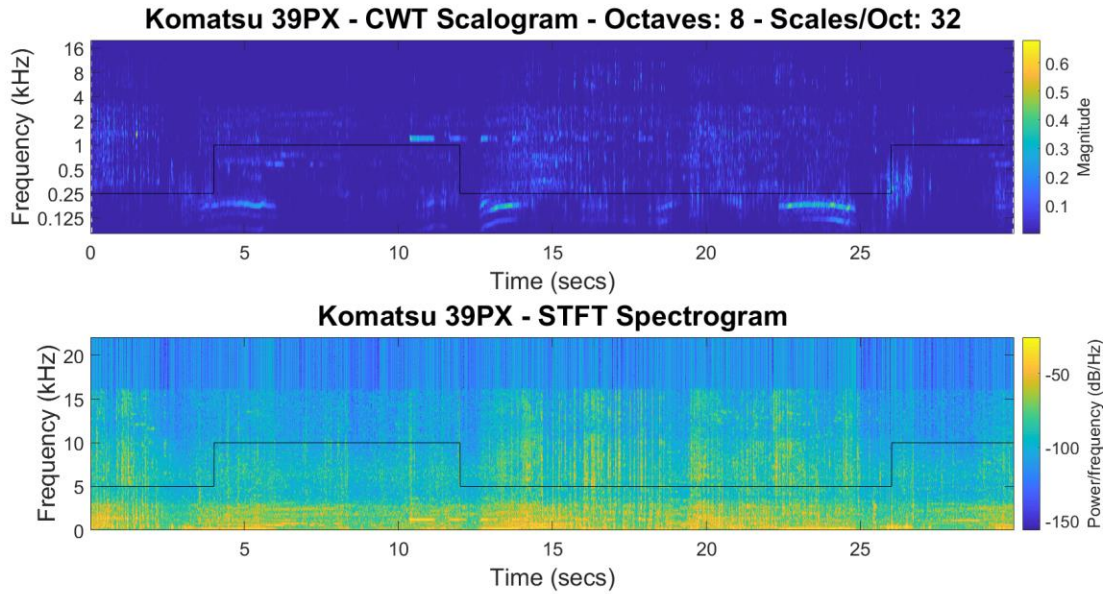
Figure 4.18: Komatsu 39PX– CWT 8/32 scalogram vs. STFT spectrogram comprison.

The confusion matrix true positives for major activities (Act 1) and minor activities (Act 2) are summarized in Table 4.2. By careful observation, it can be determined that the CWT (8 octaves and 32 scales per octave) consistently provides better results than the STFT.

Table 4.2: CWT 8/32 vs. STFT true positive classification accuracy comparison.

	CWT 8/32		STFT	
	Act 1	Act 2	Act 1	Act 2
JD 700J	82.19%	97.36%	93.44%	87.53%
JD 670G	92.86%	78.73%	79.31%	79.28%
JCB 3CX	89.01%	68.21%	91.44%	54.62%
Komatsu PC200	62.05%	79.45%	62.18%	72.75%
Komatsu 39PX	89.20%	90.18%	85.53%	82.19%

In general, it can be observed that CWT scalograms yield a better time-frequency magnitude representation than STFT spectrograms. Additionally, it has been proved 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.

4.3　Audio Data vs. Active Sensor Data

The performance of mobile micro-electro-mechanical-systems (MEMS) accelerometers data processing framework was compared against the audio signal processing framework using recordings taken at local jobsites for the following equipment: 1) JCB 3CX backhoe excavator, 2) Komatsu 39PX dozer, 3) Caterpillar 420D backhoe excavator, and 4) John Deere 550J dozer. Each MEMS recording was submitted to denoising, time-frequency feature extraction the CWT (4 octaves and 48 scales per octave), and SVM training using 8 to 24 seconds of data for activity 1 (major activity) and 8 to 24 seconds of data for activity 2 (minor activity). Each audio recording was submitted to denoising, time-frequency feature extraction the STFT (1024 frequency points, 512-sample window, and 256 overlapped samples), and SVM training using 8 to 24 seconds of data for activity 1 and 8 to 24 seconds of data for activity 2. Finally, activity classification and filtering were performed to independent MEMS and audio data sets. These data sets were limited to being less than three minutes long due to Wi-Fi connectivity issues between the mobile device on board the heavy equipment and the laptop on site. The classified labels are shown graphically in Figures 4.19, 4.21, 4.23, and 4.25. In each figure, the upper plot shows the labels obtained using the MEMS data, the middle plot shows the labels obtained using the features extracted using the audio data, and the bottom plot shows the observed, correct labels. Major activities are represented by a high position and minor activities are represented by a low position. The MEMS CWT and the audio STFT representations of the full data sets are depicted in Figures 4.20, 4.22, 4.24, and 4.26. A black line depicting major and minor activities was plotted over the figures to provide physical significance to the frequency magnitude graphs.

The MEMS versus audio comparison graphs for the JCB 3CX backhoe excavator are shown in Figures 4.19 and 4.20. Major activities included scooping and dumping soil while the minor activity was maneuvering.

Figure 4.19: JCB 3CX – MEMS vs. Audio labeling comparison.
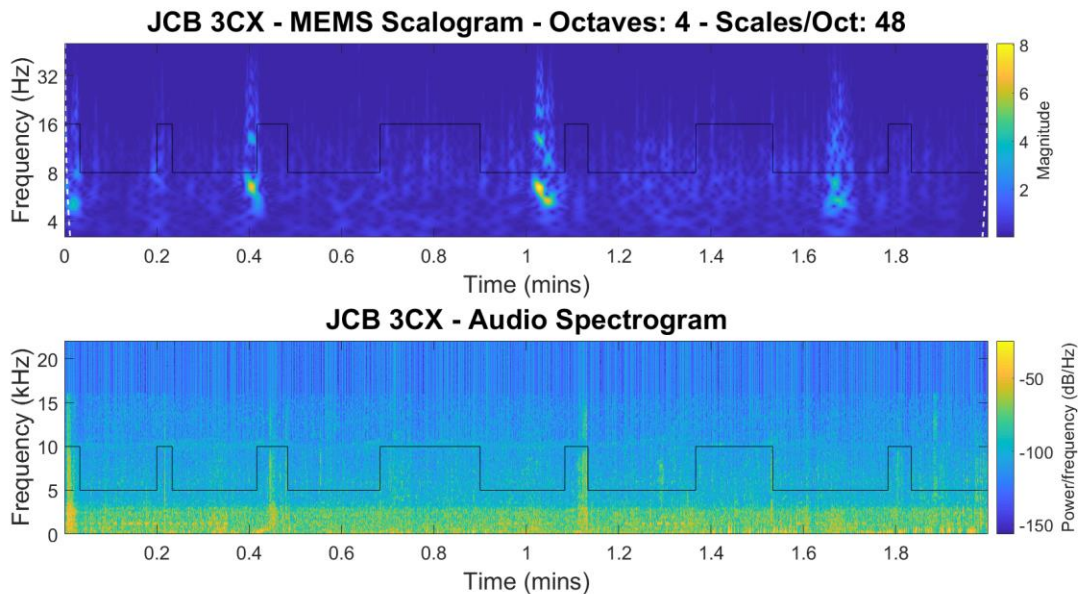
Figure 4.20: JCB 3CX– MEMS CWT 4/48 scalogram vs. Audio STFT spectrogram comprison.

The MEMS versus audio comparison graphs for the Komatsu 39PX dozer are shown in Figures 4.21 and 4.22. The major activity for this machine was pushing soil with blade while minor activities were reversing and maneuvering.

Figure 4.21: Komatsu 39PX – MEMS vs. Audio labeling comparison.
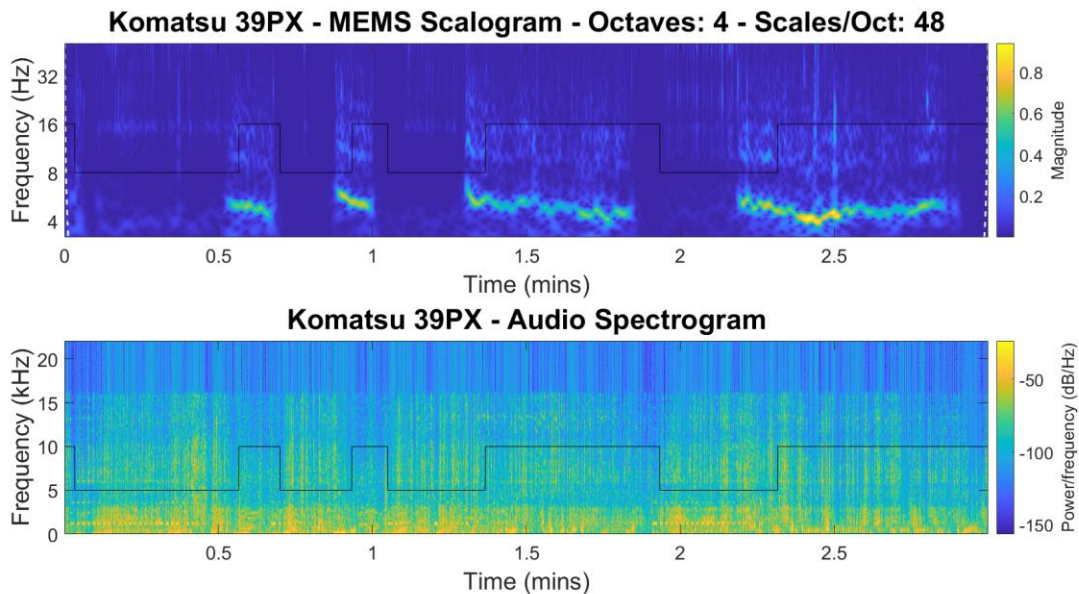
Figure 4.22: Komatsu 39PX– MEMS CWT 4/48 scalogram vs. Audio STFT spectrogram comprison.

The MEMS versus audio comparison graphs for the CAT 420D backhoe excavator are shown in Figures 4.23 and 4.24. Major activities included lifting and handling heavy structures while minor activities included maneuvering and idle times.

Figure 4.23: CAT 420D – MEMS vs. Audio labeling comparison.

Figure 4.24: CAT 420D– MEMS CWT 4/48 scalogram vs. Audio STFT spectrogram comprison.

The MEMS versus audio comparison graphs for the JD 550J dozer are shown in Figures 4.25 and 4.26. The major activity for this machine was pushing soil with blade with the minor activity was reversing.
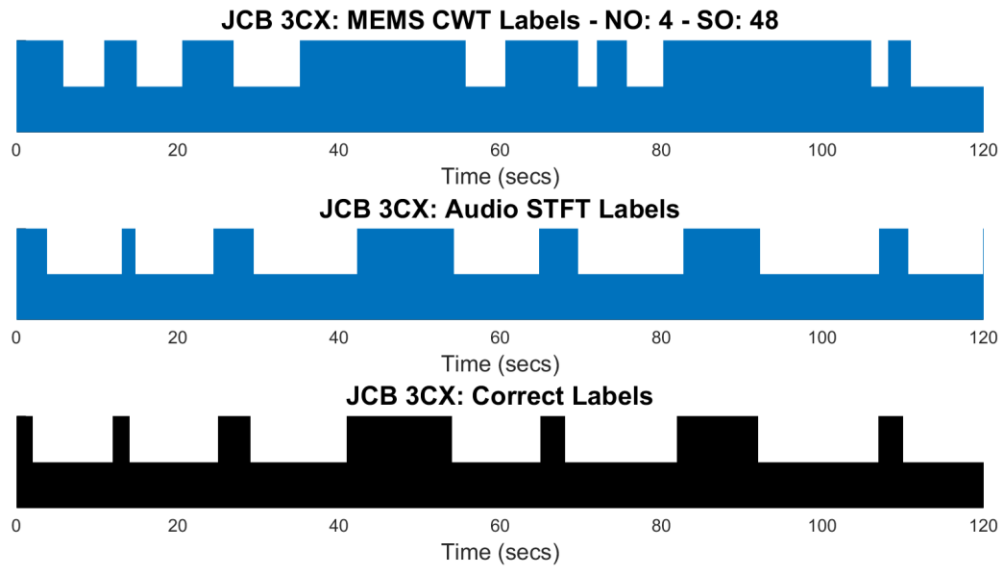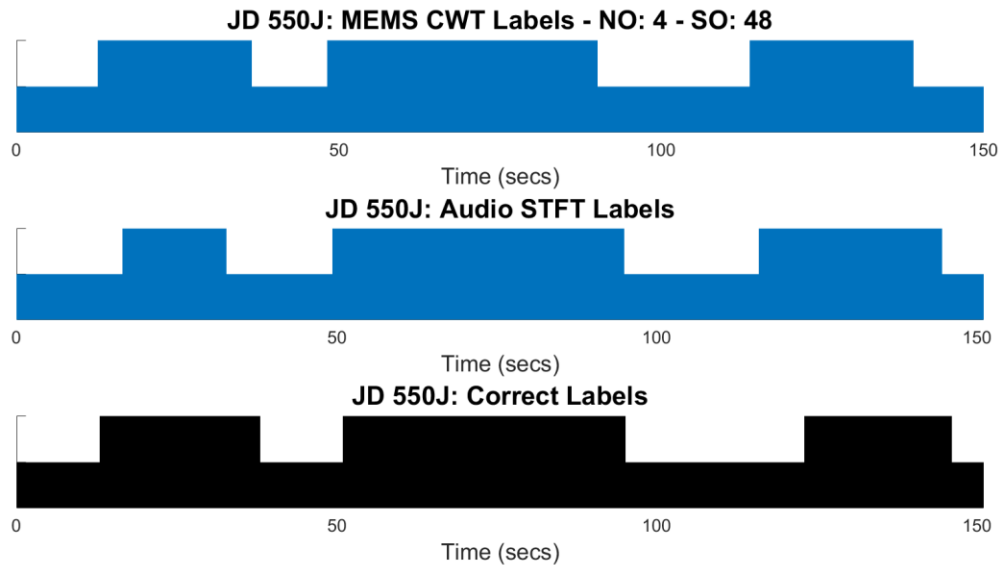

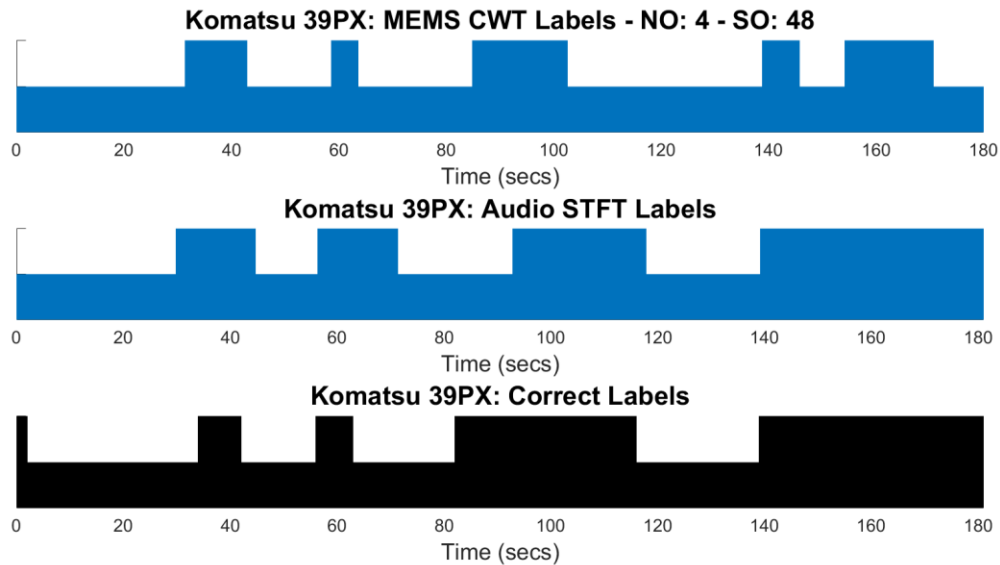Figure 4.25: JD 550J – MEMS vs. Audio labeling comparison.

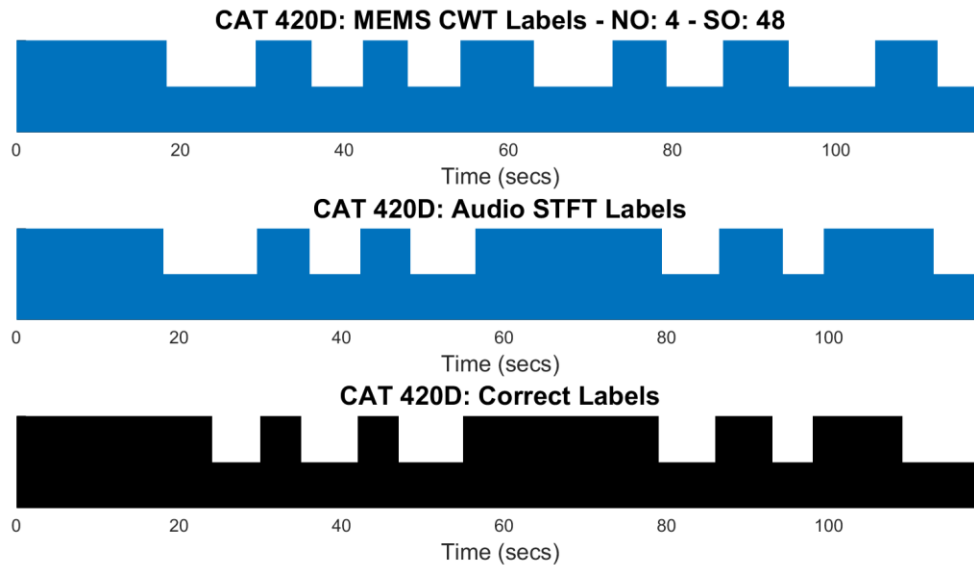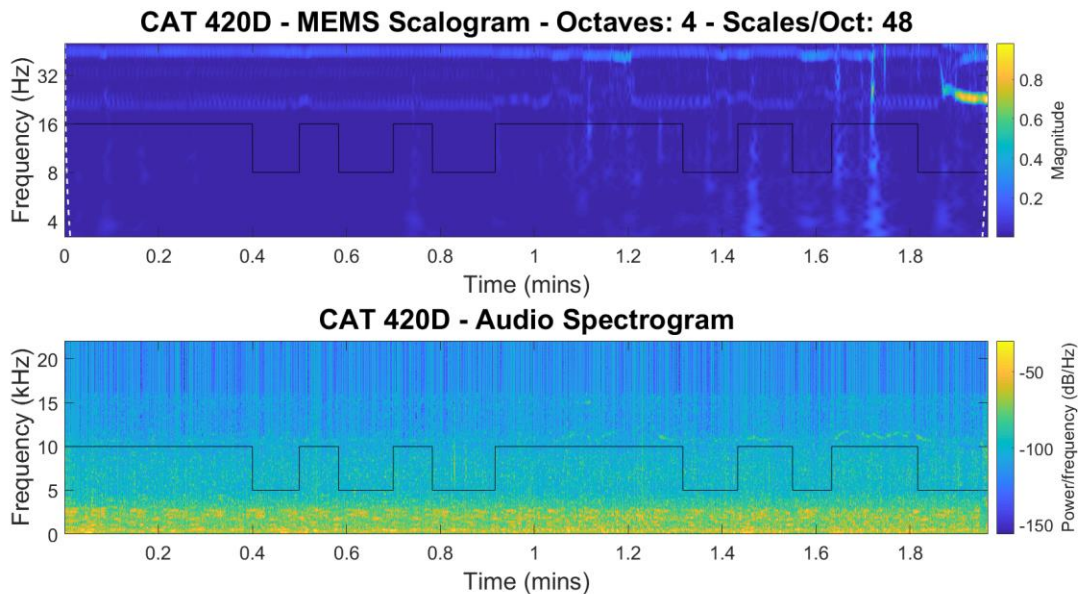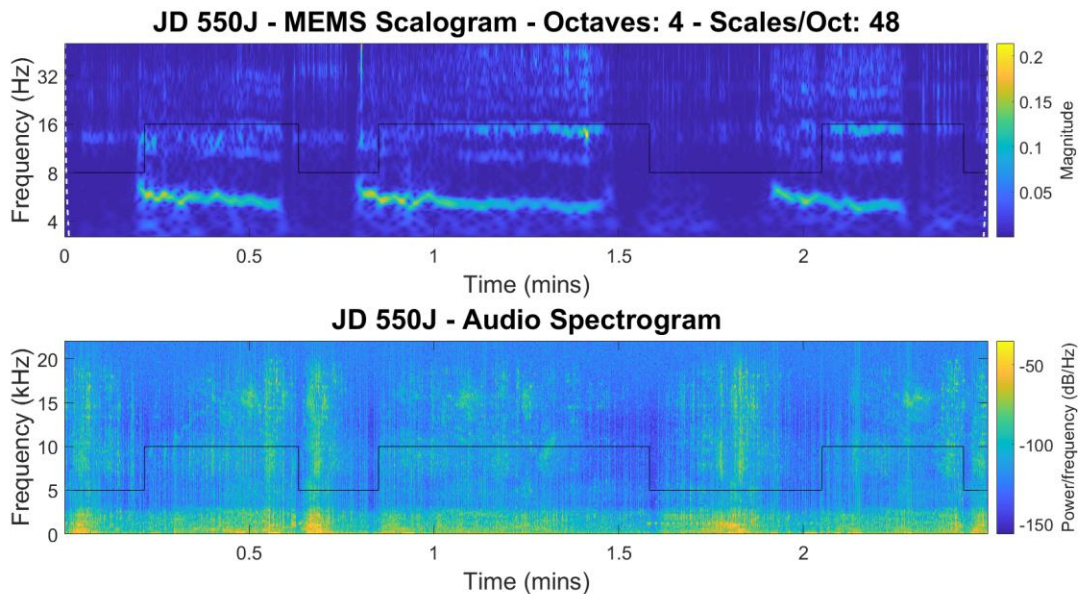Figure 4.26: JD 550J– MEMS CWT 4/48 scalogram vs. Audio STFT spectrogram comprison.

The confusion matrix true positives for major activities (Act 1) and minor activities (Act 2) are summarized in Table 4.3. By careful observation, it becomes clear that processing audio signals is better than the MEMS approach.

Table 4.3: MEMS vs. Audio true positive classification accuracy.

	MEMS		Audio	
	Act 1	Act 2	Act 1	Act 2
JCB 3CX	91.15%	47.05%	91.39%	92.39%
Komatsu 39PX	58.08%	95.03%	85.55%	80.66%
CAT 420D	70.03%	80.73%	87.27%	80.53%
JD 550J	85.43%	78.62%	88.49%	85.24%

In cases where activities were fairly simple and vibrations could be easily recognized, as with the JD 550J dozer, the accuracy achieved through the MEMS framework was very close to that achieved through the audio framework. In that particular case, there was only one activity per class, one major activity and one minor activity. In the rest of the cases, classification accuracy for the MEMS framework was unbalanced.

A notable hardware limitation about using MEMS devices connected via Wi-Fi is an apparent signal transmission lag that produces an acceleration and time label mismatch. This can be clearly observed in Figures 4.22 and 4.26, where frequency features are ahead of correct activity labels. This is very likely to be affecting classification accuracy and would be a condition to consider in case of considering further research into MEMS devices.

4.5   Audio Data and Active Sensor Data Integration

To test the potential of improving classification results by combining audio and MEMS data, a data set was processed. Per Table 4.3, major activity (Act 1) and minor activity (Act 2) classification accuracy was usually better through audio data. However, for the Komatsu 39PX

dozer, the classification accuracy for minor activities is much better using MEMS data than using audio data. Thus, such data set was selected for experimentation.

The results are presented in Figure 4.27 and Table 3.1. By careful observation, it can be noticed that a combination of audio and MEMS helps achieve midpoint on major and minor activity classification accuracy. The combined classification accuracy for Act 1 was 3.45% lower than using audio alone and the combined classification accuracy for Act 2 was 7.37% higher than using audio alone. While such results are promising, no clear conclusion can be drawn by processing a single data set.
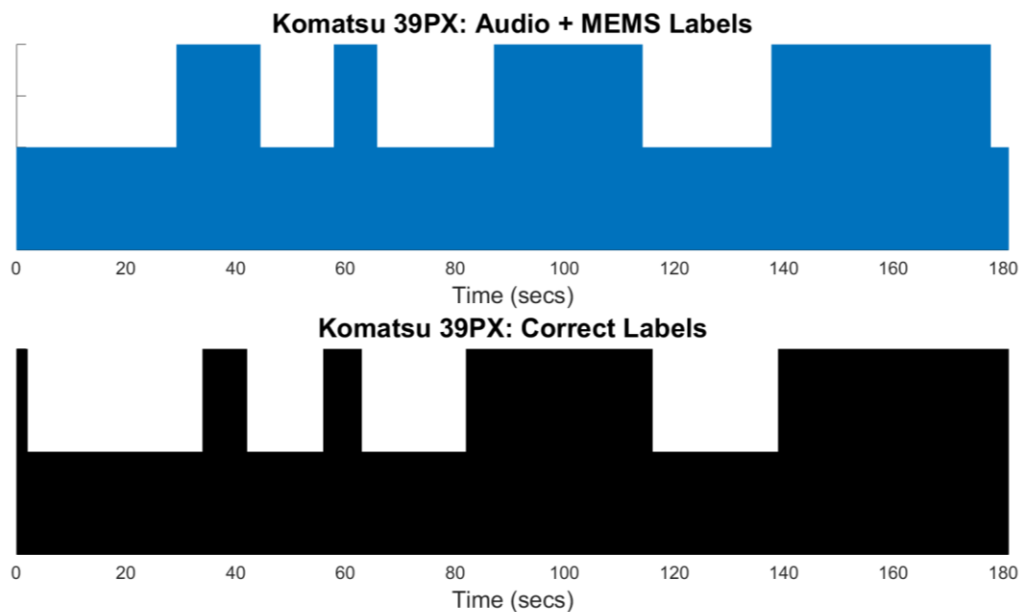

Figure 4.27: JD 550J– MEMS CWT 4/48 scalogram vs. Audio STFT spectrogram comprison.

Table 4.4: Cycle time estimation accuracy for single day analysis.

Komatsu 39PX MEMS + Audio

	Act 1	Act 2
MEMS Only	58.08%	95.03%
Audio Only	85.55%	80.66%
Combination	82.10%	88.03%

4.6   Single-Day Cycle Time Forecasting

To assess the accuracy of the cycle time estimation framework, it was initially tested with audio data for five pieces of equipment. For each machine, one audio signal was separated into two portions. The first portion was used for SVM framework training and Markov process design and an independent portion was used to test the accuracy of cycle time estimation framework. Additionally, CWT and STFT time-frequency representations were processed for preliminary evaluation. Predicted cycle times and observed cycle times are presented in Table 4.5. There, it can be observed that cycle time estimation error using the STFT is less than 7.10% and that error using the CWT is less than 5.10%. The average error obtained through the STFT and the CWT was 2.88% and 2.98%, respectively. Thus, after applying the Markov filter, the accuracy of both approaches is comparable. Nonetheless, calculating the CWT for one minute of audio recording takes about three minutes of processing. That is an important inconvenience when considering to process 30-minute audio recordings for multiple days of monitoring. Thus, STFT was preferred.

Table 4.5: Cycle time estimation accuracy for single day analysis.

Machine	Operation	Observed cycle time	Predicted cycle time (STFT)	Error (STFT)	Predicted cycle time (CWT)	Error (CWT)
JCB 3CX	Clearing	50.77 s	49.70 s	2.11%	49.90 s	1.71%
CAT 320E	Excavating	9.30 s	9.96 s	7.10%	9.02 s	3.01%
JD 700J	Grading	50.22 s	50.56 s	0.68%	52.78 s	5.10%
CAT 320D	Excavating	12.51 s	12.68 s	1.36%	13.10 s	4.72%
JD 670G	Grading	65.47 s	63.39 s	3.18%	65.71 s	0.37%

An example of labeled audio signal for a John Deere 670G motor grader while leveling ground is depicted in Figure 4.28. Given that the STFT was used for multiple-day analysis, it was used for generating these labels. The top graph in the figure shows the predicted labels directly after the SVM framework, the middle graph in the figure shows the predicted labels after the Markov chain filter, and bottom graph shows correct, manually labeled activities. A high position

represents a major activity and a low position represents a minor activity. Directly after the SVM framework, the predicted cycle time is 0.2 seconds. This is not even close to the observed cycle time because of minuscule oscillations that the window filter fails to eliminate. These oscillations become even more noticeable when zooming into the image (Figure 4.29). From the predicted sequence after the Markov filter, an average cycle time of 65.47 seconds has been estimated. The observed average cycle time was 63.39 seconds, which yields an estimation error of 3.18%. Thus, the Markov filter generates better higher order labels than window filtering alone.
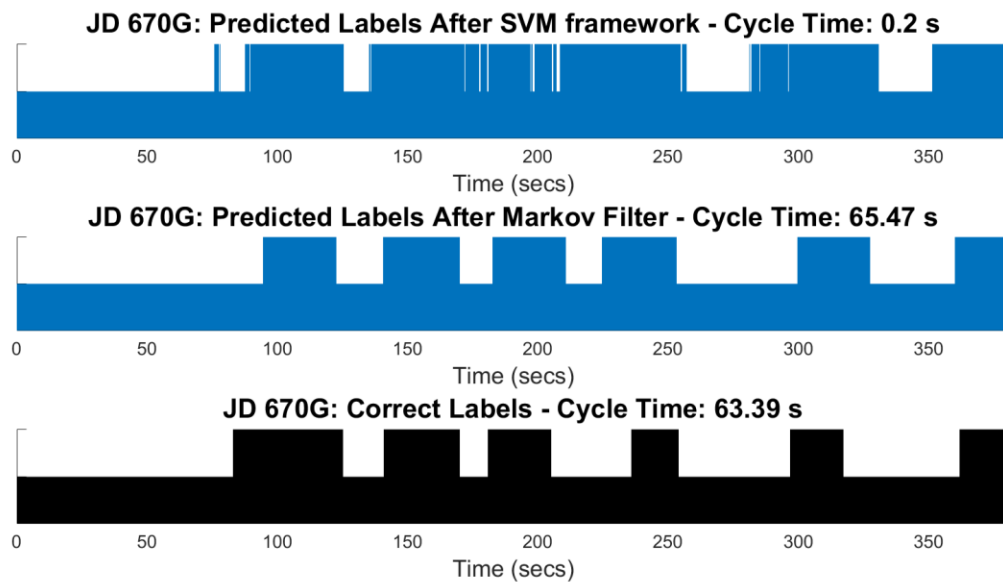
Figure 4.28: Labeled activities for a JD 670G grader leveling ground.
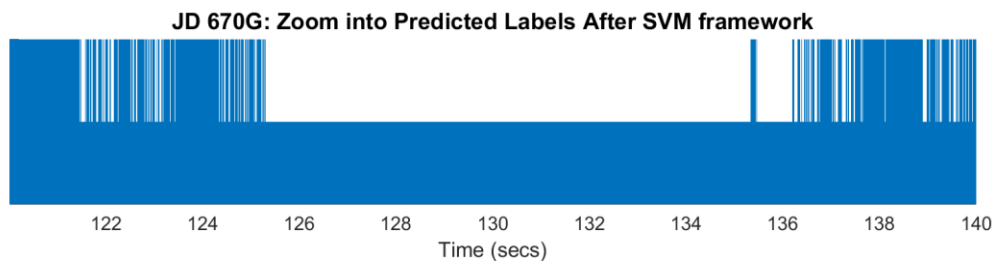
Figure 4.29: Zoom into seconds 120 to 140 of SVM-labeled activities.

4.7   Multiple-Day Cycle Time Forecasting

Typical machine operation is carried under varying work conditions (e.g., location accessibility, weather conditions, and operator skill). Thus, additional experimentation was performed to evaluate the accuracy of the cycle estimation framework over several days of operation. A Komatsu PC200 excavator was monitored during 5 days while crushing and moving demolition material (Figure 4.30). For each day, a 12 to 30-minute audio signal was processed and the estimated cycle time was compared against the observed cycle time obtained from manually labeled activities. The predicted average cycle time and observed average cycle time for each day are plotted in Figure 4.31. It can be observed that estimation error only exceeded 10% on day 4, as observed in Figure 4.32. For that specific occasion, operation was slightly different. Instead of loading the truck on a static position, the operation included maneuvering to carry material. Hence, cycle time was greater than usual.
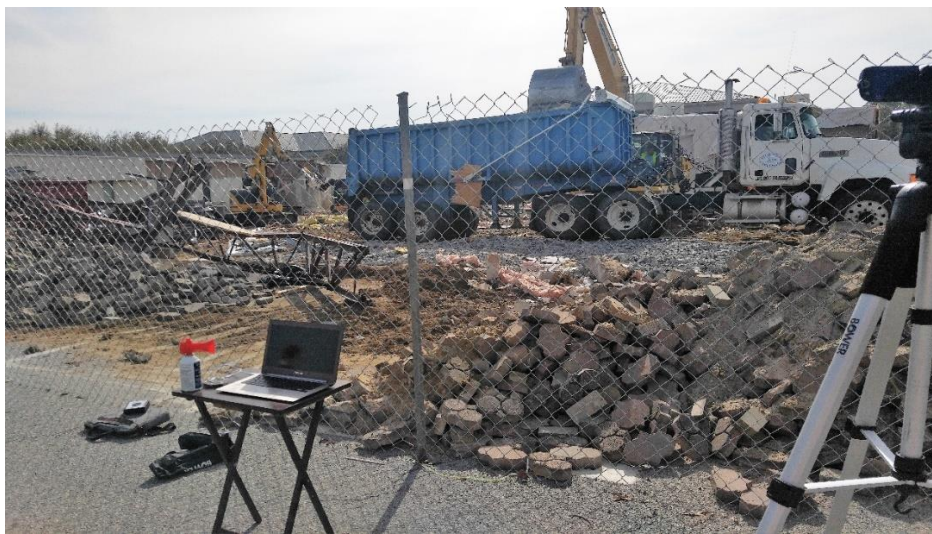
Figure 4.30: Simultaneous audio and video recording.
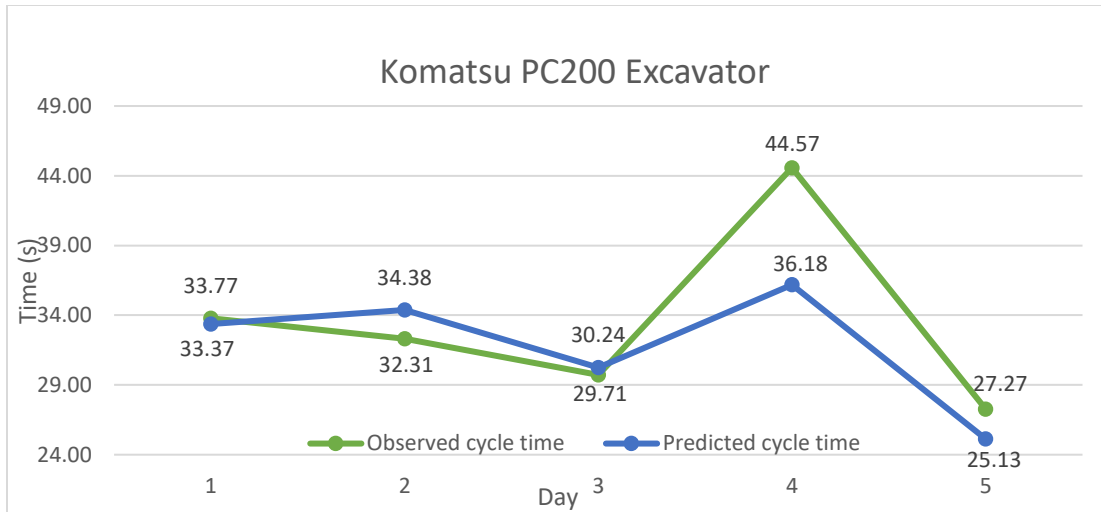
Figure 4.31: Komatsu PC200 - Observed cycle time vs. predicted cycle time.
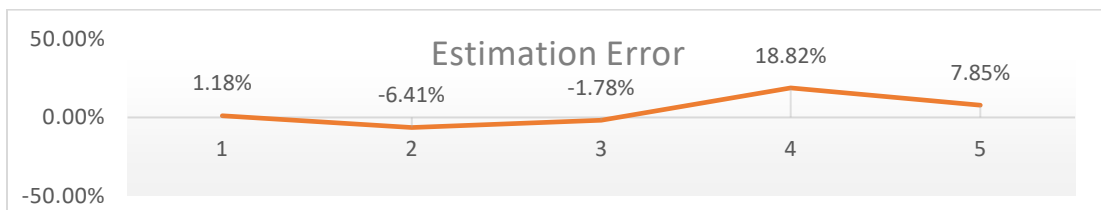
Figure 4.32: Komatsu PC200 - Cycle time estimation error.

Additionally, a JD 50G backhoe was monitored during 2 days showing cycle time estimation error of less than 10%, as shown in Table 4.6. Thus, it be concluded that a robust cycle time estimation model can be achieved through audio signal analysis and the inclusion of statistical information.

Table 4.6: Cycle time estimation accuracy for multiple day analysis.

Machine/ Activity	Description	Day 1	Day 2	Day 3	Day 4	Day 5
Komatsu PC200/ Excavating	Observed cycle time	33.77 s	32.31 s	30.24 s	44.57 s	27.27 s
	Predicted cycle time	33.37 s	34.38 s	29.71 s	36.18 s	25.13 s
	Error	1.18%	6.41%	1.75%	18.82%	7.85%
JD 50G Backhoe/ Clearing	Observed cycle time	14.54 s	11.05 s	-	-	-
	Predicted cycle time	14.92 s	10.21 s	-	-	-
	Error	2.61%	7.62%	-	-	-

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1   Conclusion

This study served to achieve a milestone toward an automated system for construction heavy equipment cycle time estimation while operating on real-world conditions. Accurate cycle time estimation is crucial because it is the basis for a real-time productivity estimation system. During the process, key hardware and software requirements were compared and the following was determined:

- Regarding audio frequency feature extraction approach for classifier training, better results were achieved through the continuous wavelet transform (CWT) employing a bump wavelet with 8 octaves and 32 scales per octave than by employing the same wavelet with 10 octaves and 24 scales per octave.

- When comparing the short-time Fourier transform versus the CWT for classifier training, better accuracy was achieved through the bump CWT with 8 octaves and 32 scales per octave. However, the CWT is much more computationally expensive and may become impractical for large audio files.

- Regarding hardware for jobsite data acquisition, better activity classification accuracy was achieved through audio signal processing than by active sensor data processing (MEMS accelerometers). However, more data sets must be processed to make definitive remarks.

- Regarding audio and active sensor data combination, there is a potential of improving audio labeling accuracy through such approach. However, more experimentation is necessary.

- Regarding audio frequency feature extraction approach for cycle time estimation, comparable results were achieved using the CWT and the STFT after applying the Markov filter.

Thus, audio signal features were extracted via the STFT and processed through the Markov filter for cycle time estimation from on-site recordings. By such method, cycle time was estimated for up to five days of single-machine operation with an accuracy over 81%.

5.2   Recommendations for Future Work

Future work could aim onto testing and improving the robustness of the current approach by focusing on the following aspects:

- Evaluating the cycle time estimation framework for more construction equipment during multiple days of operation.

- Dividing activities into sub-activities. That would expand the capabilities of the current framework, which is limited to monitoring one single action per model (e.g., excavating, truck loading, compacting, grading).

- Evaluating other machine learning algorithms beside support vector machines.

- Using more capable hardware that would allow for more experimentation with active sensors and frequency feature extraction techniques.

- Adapting the activity identification and cycle time estimation frameworks for real-time implementation.

REFERENCES

Abdoli, Mansour, and F. Fred Choobineh. 2004. "Empirical Bayes forecasting methods for job flow times." *IIE Transactions* 635-649. doi:10.1080/07408170590948495.

Adeli, M. Mahdavi, A. Deylami, M. Banazadeh, and M.M. Alinia. 2011. "A Bayesian approach to construction of probabilistic seismic models for steel moment-resisting frames." *Scientia Iranica* 855-894. doi:10.1016/j.scient.2011.07.019.

Ahn, Changbum, SangHyun Lee, and Feniosky Peña-Mora. 2014. "Application of low-cost cccelerometers for measuring the operational efficiency of a construction equipment fleet." *Journal of Computing in Civil Engineering* (American Society of Civil Engineers) 0401404201-11. doi:10.1061/9780784412329.189.

—. 2012. "Monitoring system for operational efficiency and environmental performance of construction operations using vibration signal analysis." *Construction Research Congress.* 1879-1888. doi:10.1061/9780784412329.189.

Akhavian, Reza, and Amir Behzadan. 2014. "Construction activity recognition for simulation input modeling using machine learning classifiers." Edited by S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller A. Tolk. *2014 Winter Simulation Conference.* Orlando: IEEE. 3296-3307. doi:10.1016/j.aei.2015.03.001.

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

Bengtsson, Marcus, Erik Olsson, Peter Funk, and Mats Jackson. 2004. "Technical design of condition based maintenance system—A case study using sound analysis and case-based reasoning." *Research Gate.* January. https://www.researchgate.net/publication/228697170.

Bügler, Maximilian, Ogunmakin Gbolabo, Jochen Teizer, Patricio Vela, and André Borrmann. 2014. "A comprehensive methodology for vision-based progress and activity estimation of excavation processes for productivity assessment." *EG-ICE Workshop on Intelligent Computing in Engineering.* Cardiff. 1-10.

Caterpillar. 2017. "Caterpillar Performance Handbook." Peoria, Illinois, January.

Chen, Christina Y.J., Edward I. George, and Valerie Tardif. 2001. "A Bayesian model of cycle time prediction." *IIE Transactions* 921-930.

Cheng, Chieh-Feng, Abbas Rashidi, Mark A. Davenport, and David V. Anderson. 2017. "Activity analysis of construction equipment using audio signals and Support Vector Machines." *Automation in Construction* (American Society of Civil Engineers) 81: 1-14. doi:10/1016/j.autcon.2017.06.005.

—. 2016. "Audio signal processing for activity recognition of construction heavy equipment." *33rd International Symposium on Automation and Robotics in Construction.* American Society of Civil Engineers. 1-8.

Cheng, Chieh-Feng, Abbas Rashidi, Mark A. Davenport, David V. Anderson, and Chris A. Sabillon. 2017. "Acoustical modeling of construction jobsites: hardware and software requirements." *ASCE International Workshop on Computing in Civil Engineering 2017.* Seattle: American Society of Civil Engineers. 352-359.

Cho, Chunhee, 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. doi:10.1061/9780784480847.041.

Construction Industry Institute. 2014. *Industrial Engineering/Manufacturing Techniques for Enhancing Construction Project Performance.* Research Summary, Austin: The University of Texas at Austin.

Dongarra, Jack, and Francis Sullivan. 2000. "Guest Editors Introduction to the top 10 algorithms." *Computing in Science and Engineering* (IEEE) 22-23. doi:10.1109/MCISE.2000.814652.

Golparvar-Fard, Mani, Arsalan Heydarian, and Juan Carlos Niebles. 2013. "Vision-Based Action Recognition of Earthmoving Equipment Using Spatio-Temporal Features and Support Vector Machine Classifiers." *Advanced Engineering Informatics* 652–663. doi:10.1016/j.aei.2013.09.001.

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.

Halpin, D. W., and L. S. Riggs. 1992. *Planning and Analysis of Construction Operations.* John Wiley & Sons, Inc.

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.

Intergraph Corporation. 2012. "Factors Affecting Construction Labor Productivity." September. Accessed September 20, 2017. http://bit.ly/2fURIrH.

Khosrowpour, A., J. C. Nieblesb, and M. Golparvar-Fard. 2014. "Vision-based workface assessment using depth images for activity analysis of interior construction operations." *Automation in Construction* 48: 74-87. doi:10.1016/j.autcon.2014.08.003.

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.* Montreal: International Association for Automation and Robotics in Construction. 661-668.

Kowalczyk, Alexandre. 2015. *SVM Tutorial: Understanding the Math.* June 8. Accessed September 20, 2017. http://bit.ly/2xqIQoP.

Lean Construction Intitute. 2017. *What is lean design and construction?* Accessed July 12, 2017. http://bit.ly/2vo2WhV.

Luo, Xiaochun, Heng Li, Fei Dai, Dongping Cao, Xincong Yang, and Hongling Gou. 2016. "A hierarchical Bayesian model of workers' responses to proximity warnings of construction safety hazards: towards constant review of safety risk control measures." *Journal of Construction Engineering and Management.* doi:10.1061/(ASCE)CO.1943-7862.0001277.

Lutz, James D., and Daniel W. Halpin. 1992. "Analyzing linear construction operations using simulation and line of balance." *Transportation Research Record 1351* 48-56.

MathWorks, Inc. 2017-a. *Time-Frequency Analysis with the Continuous Wavelet Transform.* Accessed October 2017. http://bit.ly/2yyMM7j.

—. 2017-b. *Understanding Wavelets, Part 2: Types of Wavelet Transforms.* Accessed October 31, 2017. http://bit.ly/2lwbVK5.

Naik, R. P. 2009. *Ultrasonic: A new method for condition monitoring.* Raipur, National Thermal Power Corporation . 10 12. Accessed 03 26, 2016. http://bit.ly/2g1ixxU.

National Instruments. 2016. *What Determines if a Transducer Is Active or Passive?* March 12. Accessed September 30, 2016. http://bit.ly/2fV3Qsx.

Navon, Ronie. 2005. "Automated project performance control of construction projects." *Automation in Construction.* 467-476. doi:10.1016/j.autcon.2004.09.006.

OpenCV. 2016. *Introduction to Support Vector Machines.* October 15. Accessed October 16, 2016. http://bit.ly/2xszZON.

Pang, Hong, Cheng Zhang, and Amin Hammad. 2006. "Sensitivity analysis of construction simulation using Cell-Devs and MycroCyclone." *Winter Simulation Conference.* 2021-2028. doi:10.1109/WSC.2006.322989.

Peurifoy, Robert, Clifford J. Schexnayder, Aviad Shapira, and Robert Schmitt. 2010. *Construction Planning, Equipment, and Methods.* 8th. McGraw-Hill.

Project Management Institute, Inc. 2016. *What is Project Management?* Accessed September 2016. http://bit.ly/2y9ITnI.

Rangachari, Sundarrajan, and Philipos C. Loizou. 2006. "A noise-estimation algorithm for highly non-stationary environments." *Speech Communication 48* 220-231. doi:10.1061/(ASCE)CO.1943-7862.0000652.

Rashidi, Abbas. 2015. "Audio-based spatio-temporal construction operations monitoring framework with real-time responsiveness." Project Description, Statesboro, 1-15.

Rezazadeh, Azar Ehsan. 2013. *Computer vision-based solution to monitor earth material loading activities.* PhD Dissertation, Department of Civil Engineering, University of Toronto.

Rezazadeh, Azar Ehsan, Sven Dickinson, and Brenda McCabe. 2013. "Server-customer interaction tracker: A computer vision-based system to estimate dirt loading cycles." *Journal of Construction Engineering and Management.* doi:10.1061/(ASCE)CO.1943-7862.0000652, 785-79.

Operations Using Audio Signals and a Bayesian Approach." *Construction Research Congress.* New Orleans: American Society of Civil Engineers.

Semaan, Nabil. 2016. "Stochastic productivity analysis of ready mix concrete batch plant in Kfarshima, Lebanon." *International Journal of Science, Environment and Technology* 5 (1): 7-16.

Shen, Chien-wen. 2008. "Bayesian estimation of defect inspection cycle time in TFT-LCD module assembly process." *International MultiConference of Engineers and Computer Scientists.* Hong Kong.

Tajeen, H., and Z. Zhu. 2014. "Image dataset development for measuring construction equipment recognition performance." *Automation in Construction* 1-10. doi:doi: 10.1016/j.autcon.2014.07.006.

Teizer, J., B. Allread, C. Fullerton, and J. Hinze. 2010. "Autonomous pro-active real-time construction worker and equipment operator proximity safety alert system." *Automation in Construction* 19: 630–640.

Torrent, D.G., and C. H. Caldas. 2009. "Methodology for automating the identification and localization of construction components on industrial projects." *Journal of Computing in Civil Engineering* (American Society of Civil Engineers) 23 (1): 1-13. doi:10.1061/(ASCE)0887-3801.

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

Yamaha. 2016. *Which types of Microphone Are Used with PA systems?* April 2. http://www.yamahaproaudio.com/global/en/training_support/selftraining/pa_guide_beginner/microphone/.

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.

APPENDIX

List of Related Publications

Cheng, Chieh-Feng, Abbas Rashidi, Mark A. Davenport, and David V. Anderson. 2017. "Activity analysis of construction equipment using audio signals and Support Vector Machines." *Automation in Construction* (American Society of Civil Engineers) 81: 1-14. doi:10/1016/j.autcon.2017.06.005.

—. 2016. "Audio signal processing for activity recognition of construction heavy equipment." *33rd International Symposium on Automation and Robotics in Construction.* American Society of Civil Engineers. 1-8.

Cheng, Chieh-Feng, Abbas Rashidi, Mark A. Davenport, David V. Anderson, and Chris A. Sabillon. 2017. "Acoustical modeling of construction jobsites: hardware and software requirements." *ASCE International Workshop on Computing in Civil Engineering 2017.* Seattle: American Society of Civil Engineers. 352-359.

Work in Progress

Cheng, Chieh-Feng, Abbas Rashidi, Mark A. Davenport, David V. Anderson, and Chris A. Sabillon. 2018. "Software and hardware requirements for audio-based analysis of contruction operations." *Journal of Computing in Civil Engineering* (American Society of Civil Engineers).

Sabillon, Chris A., Abbas Rashidi, Biswanath Samanta, Chieh-Feng Cheng, Mark A. Davenport, and David V. Anderson. 2018. "A Productivity Forecasting System for Construction Cyclic Operations Using Audio Signals and a Bayesian Approach." *Construction Research Congress.* New Orleans: American Society of Civil Engineers.