ICCEPM 2022
The 9th International Conference on Construction Engineering and Project Management
Jun. 20-23, 2022, Las Vegas, NV, USA

Collision Hazards Detection for Construction Workers Safety Using Equipment Sound Data

Kehinde Elelu[1]*, Tuyen Le[2], Chau Le[3]

[1] *The Glenn Civil Engineering Department, Clemson University, 131 Lowry Hall, Clemson, SC 29634, USA,* E-mail address: kelelu@clemson.edu
[2] *The Glenn Civil Engineering Department, Clemson University, 131 Lowry Hall, Clemson, SC 29634, USA,* E-mail address: tuyenl@clemson.edu
[3] *Department of Civil, Construction and Environmental Engineering, North Dakota State University, 1410 14th Avenue North, Fargo, ND 58102, USA,* E-mail address: chau.le@ndsu.edu

Abstract: Construction workers experience a high rate of fatal incidents from mobile equipment in the industry. One of the major causes is the decline in the acoustic condition of workers due to the constant exposure to construction noise. Previous studies have proposed various ways in which audio sensing and machine learning techniques can be used to track equipment's movement on the construction site but not on the audibility of safety signals. This study develops a novel framework to help automate safety surveillance in the construction site. This is done by detecting the audio sound at a different signal-to-noise ratio of -10db, -5db, 0db, 5db, and 10db to notify the worker of imminent dangers of mobile equipment. The scope of this study is focused on developing a signal processing model to help improve the audible sense of mobile equipment for workers. This study includes three-phase: (a) collect audio data of construction equipment, (b) develop a novel audio-based machine learning model for automated detection of collision hazards to be integrated into intelligent hearing protection devices, and (c) conduct field experiments to investigate the system' efficiency and latency. The outcomes showed that the proposed model detects equipment correctly and can timely notify the workers of hazardous situations.

Keywords: Construction Heavy Equipment, Convolutional Neural Networks, Autonomous Sound Surveillance, Construction Safety.

1. INTRODUCTION

According to the Occupational Safety and Health Administration (OSHA), the annual fatality rate in the construction industry in the United States is relatively high compared to that in other industrial sectors [1]. Struck-by equipment or vehicles is one leading cause of construction-related deaths after fall [2], mainly due to the proximity between construction workers and heavy mobile equipment [3]. The critical factor leading to collisions was reported as the decline in auditory situational awareness of construction workers due to the hearing loss [4] and the complicated nature of construction noises [5]. Therefore, a novel audio-based technique that can augment the audible sense of workers is crucial to improving safety performance.

Advanced computational techniques in auditory signal processing for hazard detection are motivated by strong acoustic emissions from hazardous situations. It is possible to extract much useful information from sounds at job sites. Mobile construction equipment often produces unique sound patterns while performing certain activities [6], [7]. However, acoustic events are typically complicated by heterogeneous sound types generated from diverse construction equipment operations, including static equipment and hand tools [8], [9]. Therefore, it is useful to distinguish between acoustic events of mobile equipment associated with collision hazards and acoustic events of stationary equipment. The auditory surveillance of potential struck-by vehicle accidents caused by mobile equipment would significantly improve construction safety. However, sound sensing for safety in the construction field has received little attention from the academic community. A majority number of related studies were focused on tracking various construction equipment activities to reduce operating costs and identifying working and operation activities [6], [7], [10], [11]. No studies have been conducted to help workers recognize important signals buried in background noises. To address the gap, this study aimed to determine whether the proposed technology improves construction workers' safety by augmenting their ability to hear important sounds related to mobile equipment in a noisy environment. This was done by identifying and characterizing distinctive features of acoustic safety cues associated with equipment-related hazards that require quick and effective responses from construction workers. The studies also developed a proof-of-concept prototype of a sound detection device that can recognize and timely provide the end-user with reliable warnings of potential moving equipment.

2. METHODOLOGY

The overall process includes three main steps: (1) collected and labeled acoustic signals as abnormal and normal types which were mixed at different signal-to-noise ratios for testing purposes, (2) extracted acoustic features using the Fast Fourier Transform (FFT) function, and (3) trained a CNN model using the labeled data to detect acoustic events.

2.1. Dataset preparation

The data preparation stage defines the set of events the system should recognize for the scope of the study, and the audio files of the dataset prepared for this research include two sources: 1) audiotapes extracted from videos downloaded from publicly available audio repositories and 2) sound recorded from construction sites of our industry partners.

Table 1. Number of original examples in each subset of data

Abnormal group (Mobile equipment)			Normal group (Stationary equipment)		
Type	Total duration (s)	Number of audio files.	Type	Total duration (s)	Number of audio files.
Bulldozer	60	20	Concrete pumper	60	20
Compactor roller	60	20	Hammer	60	20
Crane	60	20	Pile driver	60	20
Excavator	60	20	Pneumatic breaker	60	20
Forklift	60	20	Pneumatic tamper	60	20
Front end loader	60	20	Saw	60	20
Grader	60	20	Steel welding	60	20
Scraper	60	20			
Water truck	60	20			
Total	540	180		420	140

The recordings were made so that each includes only one single sound source (no overlap with other background signals) with high quality and converted into WAV format at 16 kHz sampling rate, 16-bit depth, and mono channel. The authors manually annotated the collected data with the following labels: normal and abnormal. The normal label refers to sounds of stationary equipment that are not associated with collision hazards while the abnormal category is sounds generated from mobile equipment. The non-overlap dataset is summarized in Table 1.

To generate audio examples that include concurrent sounds for testing purposes, the abnormal signals were mixed with the normal sounds at different signal-to-noise (SNR). SRN is a ratio representing how large the signal level is compared to the noise level, and the unit is in dB (decibel). A signal is an abnormal sound that needs to be detected, while noise refers to unimportant sounds of stationary equipment. The higher the SNR is, the higher the signal's amplitude is relative to that of the noise. The SNR can be calculated by the following formula:

$$SNR_{dB} = 20log_{10}\frac{A_{signal}}{A_{noise}} \tag{1}$$

This sound mixing process generates a new dataset of 44,800 audio files. Each of these is a mixture of two distinguished equipment types: mobile equipment mixed with stationary equipment or stationary equipment mixed with stationary equipment. The new audio files that include one sound from the mobile equipment group are considered abnormal, such as an excavator mixed with a hammer. The audio files do not consist of any sound from mobile equipment considered normal. The mixtures were created at different SNRs (-10dB, -5dB, 0dB, 5dB and 10dB) which was used for testing.

2.2 Feature extraction

After proper labeling and mixing of the dataset, the audios were then used to engineer the feature extraction stage by extracting Mel-Frequency Cepstral Coefficients (MFCCs), the most commonly used acoustic feature, to represent an acoustic signal for use in training [12], [13]. The extraction of MFCC includes the following three steps (1) framing and windowing, (2) Fast Fourier Transform calculation, and (3) Mel-Filter Bank and Discrete cosine Transform (DCT).

2.3 Model development

In this section, the authors provide the detailed method of audio signal processing, including the CNN model for abnormal and normal sounds classification using MFCC features that well represent the audio signals. The process of developing the model for detecting audio signals is discussed in the following subsections.

2.4 CNN model

After the feature extraction is completed, the CNN model is developed for sound detection with the array of the MFCC values as the input. The size of MFCC values is M x N, where M is the number of frames and N is the number of MFCCs. The deep CNN architecture employed in this study comprises four convolutional layers, followed by a max-pooling layer, a dropout layer, a flatten layer, and two fully functional layers connected layers to get the output. The output is a prediction of the class (normal and abnormal) to which audio belongs. We trained the CNN model with 80% of the samples used for training, and the remaining 20% were used for testing. The training procedure was stopped after 20 epochs. In the baseline model, the authors initially considered 20 MFCCs, 20 filters in the filter bank, and a window size of 500ms. Then, CNN models were trained with some modification of parameters. Finally, the best value of the number of

MFCCs, number of filters, and window size is applied to run the CNN model. Then, each model is tested on 5 test sets with 5 different SNR value (-10dB, -5dB, 0dB, 5dB, and 10dB).

3. RESULTS AND DISCUSSION

3.1. Computational performance

The performance of each model tested on each of five test sets is summarized in Table 2. This is a binary problem as the model develop is classifying the sound into mobile and stationary equipment type. Overall, the results show that the performance scores of most models increase when there is less background noise in the audio files. When being tested on the dataset without overlapping sounds, the model achieves an accuracy of 87.98%. This figure drops to 85.17% when background noise sounds are added to the clean signals at 10db SRN. The model's performance becomes relatively poor as noises are significantly louder than important signals. The accuracy of the models on the -10db SNR and -5db SNR are 50.63% and 56.85%, respectively.

Table 2. Comparison of model performance (frame size = 0.1s)

	Performance achieved on each test set					
Metrics	-10dB	-5dB	0dB	5dB	10dB	No mixture
Accuracy	0.5063	0.5685	0.6760	0.7785	0.8517	0.8798
Precision	0.5844	0.6466	0.7218	0.7911	0.8498	0.8779
Recall	0.5492	0.6069	0.6979	0.7920	0.8551	0.8858
F1-score	0.4697	0.5508	0.6716	0.7785	0.8507	0.8789

Figure 1. Sound classification android application

3.2. Field tests and experimental setup

To validate the applicability of the developed model in real construction sites, we build a mobile application on android devices using a tensor flow lite (a lightweight version of tensor for running machine learning models). This mobile application provides users with alerts of the occurrence of mobile equipment along with the probability that the detection is correct. The probability is the likelihood the equipment is detected, and a value close to 0 indicates the unlikelihood of that equipment type present, and a value closer to or 1 indicates certainty. Two separate experiments were conducted to evaluate the efficacy of the android application in detecting the occurrence of mobile equipment. The first experiment was conducted by testing it in the laboratory setting, while the second was done on the field. The input test signals include the sound of 16 and 4 types of equipment respectively for the two experiments. We performed three trials for each test signal. For each experiment set, the following outputs were recorded: type of equipment detected, the probability that the detection is correct, and the time it took to detect the signal. The loudness of the sound while experimenting ranges between 72 and 80db measured using a Decibel Meter iOS application. In this experiment, a sound source (among 16 types of sound) generated from a computer speaker is placed 4 meters away from the mobile device. In the field experiments, a site engineer was asked to carry the mobile device with the mobile application installed and stand at 10m and 20m away from the equipment. He recorded the average time and probability of hazard detection for three samples of each of the four equipment. The variation in distance is to check the impact of distance on the model because a closer device to a worker signifies more danger. Fig 3 shows the setup for testing the model on the construction site.

Figure 2. Experiment Setup in Controlled Environment (the loudness detector in the left, the computer, and the mobile device on the right)

Figure 3. Sound detection experiments with a front-end loader (left) and an excavator (right)

3.3. Testing results

Table 3 below shows the result of the experiments conducted in the controlled environment. The acoustic sensing application developed was used to monitor equipment classification and measure the time required to detect the important signal. The result shows that it took less than 10 seconds for the application to generate an alert. Yet, it is difficult to confirm whether this latency is sufficient for real-time hazard detection as it highly depends on many job-site factors such as the speed and the direction of a target mobile vehicle as well as the presence of barriers between the worker and the equipment. It is ideal to reduce the delay to allow the worker to have more time for responsive safety actions. This seems to be challenging for complex workplaces with excessive background noise. Because of that nature, our model requires significant computational power for processing a large amount of real-life data. Future work is needed that implements advanced pre-processing algorithms (i.e., de-noising) with less computational requirements to improve the overall performance of the proposed system, particularly reducing the detection duration.

Table 3. Results from the experiments in the laboratory

Equipment Type	Probability	Duration (sec)
Excavator	0.656	4.000
Bulldozer	0.768	5.667
Grader	0.826	5.222
Front end loader	0.627	7.333
Forklift	0.903	5.222
Compactor roller	0.882	5.111
Scraper	0.681	5.444
Water truck	0.672	8.278
Pneumatic tamper	0.413	7.667
Concrete pumper	0.532	9.667
Pile driver	0.601	3.111
Pneumatic breaker	0.486	7.778
Steel welding	0.604	9.000
Hammer	0.783	5.111
Saw	0.454	7.889

Table 4. Average Result from Site Investigation

Equipment Type	Probability		Duration (sec)	
Distance	10m	20m	10m	20m
Excavator	0.27	0.39	7.67	11.33
Frontend loader	0.92	0.53	3.33	14.00
Hammer	0.94	0.93	7.67	10.00
Saw	0.51	0.66	9.00	10.67

The results from the construction site experiments (see Table 4) generally show better performance when the worker is closer to the equipment in terms of probability and time to the first alert. The probability range among the equipment is 27%-94% and 39%-93%, respectively, when the worker is 10m and 20m further away from the test equipment. It is noted that the acoustic sensing in the field experiments was affected by surrounding equipment noise and human activities on the construction site. This could be the main cause of low probability for some types of equipment, such as the excavator (27% at 10m) and the saw (51% at 10m). Regarding the latency, the longest duration was 14 seconds when the device detected the sound of the frontend loader, which is still relatively short. Compared with the first set of experiments in the controlled environment, the performance in the field tests is significantly reduced. The controlled environment is free of exterior noise; therefore, this allows easy picking up sound features faster and higher confidence results.

3.4. Discussions

This study developed an AI model using CNN-based signal processing to enable the early detection of auditory signals related to potential collision hazards. We trained and tested various models with different signal-to-noise ratios. To reduce the number of false negatives (missing any mobile equipment that will likely cause harm to field workers), the recall evaluation metric, which measures how many observations our model correctly predicted over the total number of observations, was emphasized. Based on the findings (see Table 2), the probability of missing out on a potential hazard is not a major concern of the developed model. Also, the precision value, which indicates the mislabeling of normal background sound as a danger, is in an acceptable range. The processing capacity of smartphones obviously plays a great role in determining how fast it detects, and the quality of its built-in microphone is critical to the input sound captured. Moreover, the level of background noise greatly affects the efficacy of the device. The results of the implementation experiments indicated that the probability of true positives for the controlled experiment is much greater than those of field tests. It is because the test sounds in the field tests were greatly buried in background noises (e.g., nearby operations). Although this study proves that the proposed CNN model is a reliable technique to help detect potential collision hazards at the construction site, there are still areas to be improved for successful practical implementation. One limitation of this research is that the system could not capture the location of mobile equipment, and sound localization would help workers be aware of their position relative to the direction and distance to the hazard. Thus, localization is important to reduce false alarms for the system. Lastly, the notification zone for danger will be based on estimated figure for each type of equipment.

4. CONCLUSION

Collision hazards have posed serious threats to the safety of workers on site. The presented framework provides detection of this dangerous situation using CNN to classify normal and abnormal audio files mixed at different SNR. The performance indicates that the model yields reliable predictions with an accuracy of 88% in detecting abnormal sounds relating to collision hazards when the signals are not buried in background noises. The experiment monitors the time, probability, and type of equipment. We observed faster detection for some equipment when used, especially if it exhibits a unique sound characteristic and lower noise level. Delay in detecting some equipment is related to the device's latency due to limited computational power. TensorFlow-lite model is also cross-platform and can work on IoT devices, advantageous for future research in this or related field. This paper is expected to start a long series of research on acoustic monitoring of construction equipment and provide initial guidance for future research.

REFERENCES

[1] J. W. Hinze and J. Teizer, “Visibility-related fatalities related to construction equipment,” *Safety Science*, vol. 49, no. 5, pp. 709–718, Jun. 2011, doi: 10.1016/j.ssci.2011.01.007.
[2] Samantha Brown, William Harris, Raina D. Brooks, and Xiuwen Sue Dong, “Fatal Injury Trends in the Construction Industry,” 2021.
[3] E. D. Marks and J. Teizer, “Method for testing proximity detection and alert technology for safe construction equipment operation,” *Construction Management and Economics*, vol. 31, no. 6, pp. 636–646, Jun. 2013, doi: 10.1080/01446193.2013.783705.

[4] M. Tc, T. Cl, R. Rf, V. Bl, B. Dc, and E. R. Reeves, "Estimating noise-induced permanent threshold shift from audiometric shape: The ISO-1999 model Dobie RA Working in noise with a hearing loss: Perceptions from workers, supervisors and hearing conservation program managers."
[5] P. M. A. N. D. B. E. Vinnik Ekaterina AND Itskov, "Individual Differences in Sound-in-Noise Perception Are Related to the Strength of Short-Latency Neural Responses to Noise," *PLOS ONE*, vol. 6, no. 2, pp. 1–8, Dec. 2011, doi: 10.1371/journal.pone.0017266.
[6] C. F. Cheng, A. Rashidi, M. A. Davenport, and D. v. Anderson, "Activity analysis of construction equipment using audio signals and support vector machines," *Automation in Construction*, vol. 81, pp. 240–253, Sep. 2017, doi: 10.1016/j.autcon.2017.06.005.
[7] C.-F. Cheng, A. Rashidi, M. A. Davenport, and D. Anderson, "Audio Signal Processing for Activity Recognition of Construction Heavy Equipment."
[8] Y. C. Lee, M. Shariatfar, A. Rashidi, and H. W. Lee, "Evidence-driven sound detection for prenotification and identification of construction safety hazards and accidents," *Automation in Construction*, vol. 113, May 2020, doi: 10.1016/j.autcon.2020.103127.
[9] Y. Xie *et al.*, "Historical Accident and Injury Database-Driven Audio-Based Autonomous Construction Safety Surveillance."
[10] C. Sabillon, A. Rashidi, B. Samanta, C.-F. Cheng, M. A. Davenport, and D. v Anderson, "A Productivity Forecasting System for Construction Cyclic Operations Using Audio Signals and a Bayesian Approach."
[11] B. Sherafat, A. Rashidi, Y.-C. Lee, and C. R. Ahn, "A Hybrid Kinematic-Acoustic System for Automated Activity Detection of Construction Equipment," *Sensors*, vol. 19, no. 19, 2019, doi: 10.3390/s19194286.
[12] M. Cowling and R. Sitte, "Comparison of techniques for environmental sound recognition," *Pattern Recognition Letters*, vol. 24, no. 15, pp. 2895–2907, 2003, doi: https://doi.org/10.1016/S0167-8655(03)00147-8.
[13] A. J. Eronen *et al.*, "Audio-based context recognition," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 1, pp. 321–329, 2006, doi: 10.1109/TSA.2005.854103.
[14] M. Leonardo, N. Berardo, C. Alessandro, R. Luigi, and D. G. Giuseppe Martino, "Development of a Digital Twin Model for Real-Time Assessment of Collision Hazards," 2020, pp. 14–19. doi: 10.3311/ccc2020-003.