BelaySound: Detecting The Presence of Auto Belays Using A Smartphone

Woohyeong Kim
Florida State University

Abstract

Indoor rock climbing as a sport has been growing steadily in the last two decades. With its debut in the Olympics last year in Tokyo, the sport is projected to have a massive audience and participants in the near future. With the popularity, the number of climbing gyms is increasing in a unprecedented fashion. In these climbing gyms, auto belays are a big part of the business. They allow climbers to climb on high walls without a belayer, and they are extremely easy to use. However, over the last decade, auto belays have caused deaths and lawsuits simply because the climbers forgot to clip in before they started climbing.

We present BelaySound, a system that uses microphones on widely available smartphones to detect if the user is clipped in to the auto belay or not while they are climbing. When the system detects a climber who did not clip in, it will notify the staff and the climber through an alarm.

1 Introduction

Rock climbing as a sport has been growing steadily for little more than three decades now. What started off as an outdoor hobby turned into a big indoor sport in the last two decades. With the popularity of indoor climbing increasing, we are seeing a record number of climbing gyms getting built around the world. In the U.S. alone, we have almost 600 climbing gyms [2]. The number of climbing gyms around the U.S. has only been increasing since the beginning as you can see in figure 1, and with its debut at the 2021 Tokyo Summer Olympics, the sport is projected to grow even at a faster pace.

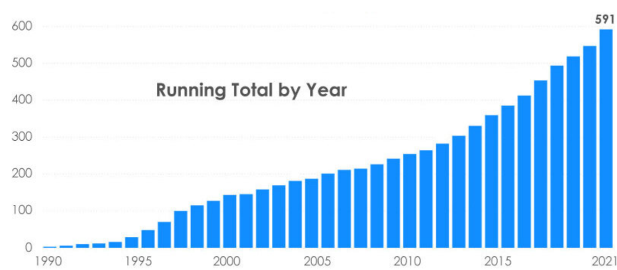

Figure 1: Number of climbing gyms in the U.S. by year

Modern climbing gyms offer two types of climbing: bouldering and rope climbing. Generally speaking, rope climbing is safer than bouldering due to multiple reasons. Firstly, rope climbing is less physically intensive. Because the number of moves in a rope route is much greater than in a boulder, the intensity tends to spread out throughout the whole rope route, making each move much less intensive than in a boulder. Secondly, you rarely hit the ground when falling on a rope route. A big portion of the climbing injuries comes from sprained ankles after a boulder fall to the ground. However, on a rope route, you almost never hit the ground since you are caught by the rope.

Even though bouldering is much more dangerous than rope climbing, most first time climbers prefer bouldering. This is because bouldering is cheaper and more approachable to beginner climbers. To climb in a rope, you much be in a harness which costs at least $100 to buy or around $8 to rent. This is in contrast to bouldering where you only need a pair of shoes. Also, rope climbing is much harder for beginner climbers to get into because you must know how to belay and have a partner who can also belay.

A nice middle ground is an auto belay. Although you still do need a harness to climb on an auto belay, you don't need to know how to belay or have a partner. Since the physical intensity is much lower and there's a very little chance of hitting the ground, an auto belay has become quite popular over bouldering among beginner climbers.

However, auto belays have faults of their own. It turns out that auto belays can get really dangerous due to user error. One of the main issues with auto belays is that the climber simply forgets to clip in before they start climbing. The auto belay manufacturers and the climbing gyms have tried to tackle this problem by having a huge, eye-catching triangle that the auto belays attach to so that the climbers must unclip the auto belay from this big triangle and attach it to themselves before climbing.

Even though this has undoubtedly helped in a way, the auto belay accidents keep happening. After this change was in place, there were multiple deaths and injuries from auto belay users. One big example is a death in Boulder, Colorado who forgot to clip in and suffered a fatal fall. Even at the local level here in Tallahassee, we've had two different unrelated incidents over a 10-month period. One suffered from multiple compression fractures in their spine, while the other noticed their mistake and had to climb all the way down.

In this work, we present BelaySound, a system designed to alert climbers and climbing gym staff when there is a climber on the wall who forgot to clip in to the auto belay. The system uses the recording of the climber (e.g. sound of their hands and feet touching the climbing holds and the wall) to detect if the climber is on the wall and to detect if there is a retracting auto belay.

The system uses two different deep neural networks: one for detecting if the climber is on the lane we are interested in, and one for detecting if the climber is clipped in or not. Also, note that the microphone is attached to the wall, not on the climber, to make sure that the system doesn't get in the way of the climbers.

In the upcoming sections, we will look at some related works (section 2) from the class followed by the main challenges we must overcome (section 3). Then, we will entail the data collection process (section 4) and the system design (section 5). The evaluation of the system (section 6) will be given next. For the last part of the paper, we'll discuss the limitations (section 7) and future works (section 8) to improve the features and the accuracy. The paper will finish with a short conclusion (section 9).

2 Related Works

There are two works from the class that are closely related to this project. In this section, we will go over each one of them in high level and how they inspired the techniques used in this paper.

2.1 SpiroSonic

SpiroSonic [3] utilizes the microphones on commonly available smartphones to carry out at-home spirometry. It has high accuracy despite having no extra cost to the user except for owning a smartphone and possibly the cost of the smartphone application. The system fills an important gap in the portable spirometer market.

Although SpiroSonic actively generates ultrasound to measure the distance from the smartphone to the user's chest, the sound processing and the use of deep neural networks are inspiring to this project.

2.2 BreathPrint

BreathPrint [1] is a similar system that uses microphones on commodity smartphones to authenticate users through their breathing patterns. The system utilizes three different breathing gestures: sniff, normal, and deep breathing. The results show a good accuracy, but the system requires a lot more improvement before it's useful in real-world scenarios.

The application of BreathPrint is quite similar to ours because the sound of auto belays lie in the lower frequency domain. Ideally, we would have liked to use Gammastone Frequency Cepstral Coefficients (GFCC) as shown in BreathPrint, but the libraries were limited. Given the short time frame of this project, we decided that Mel Frequency Ceptral Coefficients (MFCC) is sufficient for a proof-of-concept.

3 Challenges

There are multiple challenges that we must solve for BelaySound to work in a real-world scenario. However, since the scope of this project is quite limited, we entail only the challenges we must overcome for a proof-of-concept of the system.

The first is detecting which auto belay line the climber is on. Because the clip detection only works on a single auto belay line, we must be able to figure out which line the climber is on before running the clip detection algorithm. For this paper, we are only interested in the auto belay line 2. Once we infer that certain segment of the audio came from a climber on line 2, we can feed this audio segment into the clip detection algorithm.

The second is the intermittent presence of an auto belay in the sound recording. When a climber is on the wall, the auto belay does not retract continuously. This is true even when the climber moves continuously upwards. Not only that, the retraction of the auto belay depends on the pace the climber is climbing. Our system must be able to catch different retracting rates of the auto belay.

The third challenge is the amount of work that must be done to collect our data. Climbing is a highly physical activity, and collecting data can be time consuming and tiring. Also, for the first proof-of-concept, we decided that the dataset should contain no background noise or music. This makes data collection impossible during the open hours of the local climbing gym.

Another challenging aspect of the data collection process is that we must be able to segment the audio recording precisely when the climber starts climbing to when the climber has finished climbing. Because the sound of a climber can be extremely quiet, this is hard to do from a normal continuous recording of the climber relying purely on human ears.

The last challenge is collecting data for dangerous climbing, which is in our case climbing without clipping into an auto belay. Since we are trying to detect dangerous climbing through machine learning, we must present dangerous data points to the network during training. This is hard without breaking the local gym rules on dangerous behaviors.

4 Data Collection

In this section, we detail the data collection process. We will go over different methods we used to overcome the challenges that pertain to the data collection process from the previous section. Then, we will describe the datasets in more detail.

Since the time and effort are limited for a class project, we wanted to keep the scope small. One of the ways we did that is keeping our datasets clean of the background noise or music. To achieve this, the data was collected in the morning by ourselves when the gym is closed.

Risk management was an important factor in our data collection process since we must collect dangerous data (i.e. climbing without clipping in to an auto belay). We restricted our datasets to the first six feet of climbing from the ground to minimize the risk while not compromising the accuracy of the system. Another implication of this restriction in the dataset is that the alarm would go off when the climber reached around six feet. This is appropriate because we want to warn the staff and the climber before the climber reaches a life-threatening height if they are not clipped in to the auto belay.

A big part of the data collection process was being able to precisely segment each climb. This is important for the machine learning algorithms to learn the pattern accurately. Since it is quite hard for human ears to detect when a climb begins or ends from the recording, we used a specific knock patterns before and after a climb. The audio was segmented from one knock pattern to another, ensuring that the audio segment began immediately before a climb and ended shortly after with a very short delay. We also made sure none of the knocking sound was within the segment.

In machine learning, you want to expose your deep neural network to all possible patterns of data. To ensure this, we used various climbing patterns (using different climbing holds every time) and climbed with various speeds for each audio segment. This is especially important in our system because most climbers

Climb Length	Mean	Min	Max
No Clip	12.67	9.0	17.0
Clipped	12.7	10.0	18.0

Table 1: Lane Detection Dataset Statistics

Climb Length	Mean	Min	Max
Lane 1 No Clip	13.8	11.0	17.0
Lane 1 Clipped	15.0	10.0	22.0
Lane 2 No Clip	12.2	9.0	17.0
Lane 2 Clipped	12.14	10.0	14.0

Table 2: Lane Detection Dataset Statistics

have different climbing styles and skill levels that vary widely. These variations cause some climbers to use different holds and climb at different speeds.

4.1 Clip Detection Dataset

For the clip detection dataset, there are two different scenarios. The first one is a climber who climbs up to six feet with a proper use of an auto belay, and the second is the same climber without clipping into an auto belay.

We collected 30 data points for each scenario, making the whole clip detection dataset size to be 60. Detailed statistics on the clip detection dataset is shown in table 1

4.2 Lane Detection Dataset

For the lane detection dataset, there are two different scenarios on each lane. We need to make sure that we collect the same number of data points for both unclipped and clipped scenarios.

For each lane, we collected 15 data points where we climb without an auto belay and 15 data points where we climb with an auto belay. Since the scope of this project contains only two auto belay lanes, the entire lane detection dataset contains 60 data points.

Detailed statistics on the lane detection dataset is shown in table 2.

5 System Design

In this section, we will go over the details of our system design. We discuss from small design choices to different technical details that make up our system.

The first thing we must consider is the placement of the smartphone. For this project, we placed the smartphone behind the climbing wall so as to not disrupt the climbers. Also, the device was placed at around eye level height. This was an intentional design choice. We speculate that climbers will generate more noise with their feet rather than their hands, and by placing the smartphone at around five to six feet high, the sound amplitude would be the highest when the climber is at five or six feet off the ground.

In order to input the audio into our neural networks, we must extract some features first. We use 24 Mel Frequency Ceptral Coefficients (MFCC) and their deltas for every 500ms window. The beginnings of two consecutive windows are 125ms apart, so they are overlapping. Even though other sound processing systems use much smaller window sizes (30 - 50ms), we found that windows of size 500ms and higher produce the best results for our application. We speculate this is because of the intermittent presence of the auto belays in the recordings.

The two deep neural networks in our system in conjunction with each other. Lane detection algorithm first processes the input and determines the climber's lane. Then, the system invokes the clip detection network for the corresponding lane and passes on the input. This means that there will be some delays before clip detection can detect dangerous climbing, however, the delay is negligible to the speed a climber can ascend on the wall. A high level depiction of the system design is given in figure 2.

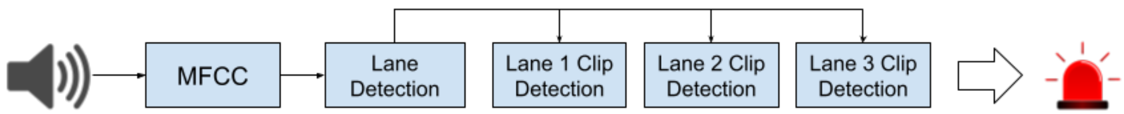

Figure 2: High level depiction of the system

Both the clip and lane detection networks have the same architecture. Each network starts with an LSTM layer followed by 7 linear layers of decreasing sizes. The LSTM layer takes in tensors of size 48 and outputs tensors of size 100, while the linear layer sizes are 100, 80, 60, 40, 20, 10, 2, respectively. We use the ReLU activation function and dropout layers with $p = 2$ for regularization. The output layer is a softmax layer since we want probabilities.

For the outputs of each network, we apply some postprocessing. Consider the output of each network to be a signal in the time domain. Then, we smooth the signal with an averaging window of size 5 across time. Now, every output is classified into class 0 or 1 depending on its amplitude. After classification, we apply a voting window of size 5 on the output. The voting window replaces the value in the middle of the window with the value that occurs most frequently within the window. The postprocessing step helps mitigate the impact of intermittent presence of the auto belay in the recording.

6 Evaluation

In this section, we will walk through the evaluation of our system. We will discuss the setup, and the performance of the system on both segmented recordings and continuous recording of multiple climbs.

6.1 Setup

For clip detection, the training dataset contains 24 clipped and 24 unclipped data points, while the test set contains 6 of each scenario. In total, the training and test sets contain 48 and 12 examples, respectively. For lane detection, there are 24 (12 clipped and 12 unclipped) data points from lane 1. Likewise, there are 24 data points from lane 2. In the test set, there are 6 (3 clipped and 3 unclipped) data points from lane 1, and the same amount of examples from lane 2. The total number of examples in the training and test sets are 48 and 12, respectively.

We assign a label (ground truth) for every MFCC window. The labels are identical for the whole audio segment (either 0 or 1). In clip detection, 0 signifies climbing with an auto belay while 1 signifies dangerous climbing. In lane detection, 0 means the first lane while 1 means the second lane.

6.2 Evaluation

The evaluation of the system simply measures the accuracy of the two neural networks. As you can see in figure 3, both networks achieve around 98% accuracy on the training sets, while they are only able to reach about 96% accuracy on the test sets.

Note that this accuracy is only achieved on the segmented audio inputs. In a real-world scenario, the microphone will be recording continuously, so the input is never segmented. This is important because the internal state of the LSTM layer will be non-zero on a continuous recording when the relevant audio segment is inputted. We will look at how the networks perform under this circumstance next.

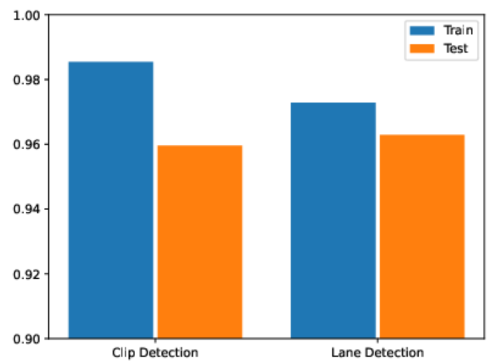

Figure 3: Evaluation of the system

6.3 Continuous Data

The evaluation on the continuous data is done qualitatively for the scope of this paper. The main reason being that we do not want to insert any audio cues for the beginning and the end of each climb. This will affect the networks in an unpredictable manner, which will interfere with our evaluation. Without the audio cues, it is extremely hard to know the exact beginning and end of each climb, which makes a quantitative evaluation difficult.

The evaluations on the continuous data for both networks are shown in figure 4 and figure 5. The x-axis is time (index of the MFCC windows) and the y-axis is the probability output from the networks. The blue curve is the output after the initial smoothing operation, while the yellow curve is after the classification and applying the voting window.

The green and red lines at the bottom and top are the ground truth values. The dotted ground truth lines in clip detection are examples from lane 1, so they shouldn't count towards the accuracy of the clip detection network. However, you can see that even when the clip detection network was only trained on lane 2 data, it's still able to infer data from nearby lanes with some accuracy.

Overall, we conclude that accuracy on continuous data needs more work. For one, the outputs of the networks at the beginning of each climb segment seem to be the region with most errors. Also, there are some segments where the networks seem quite confused.

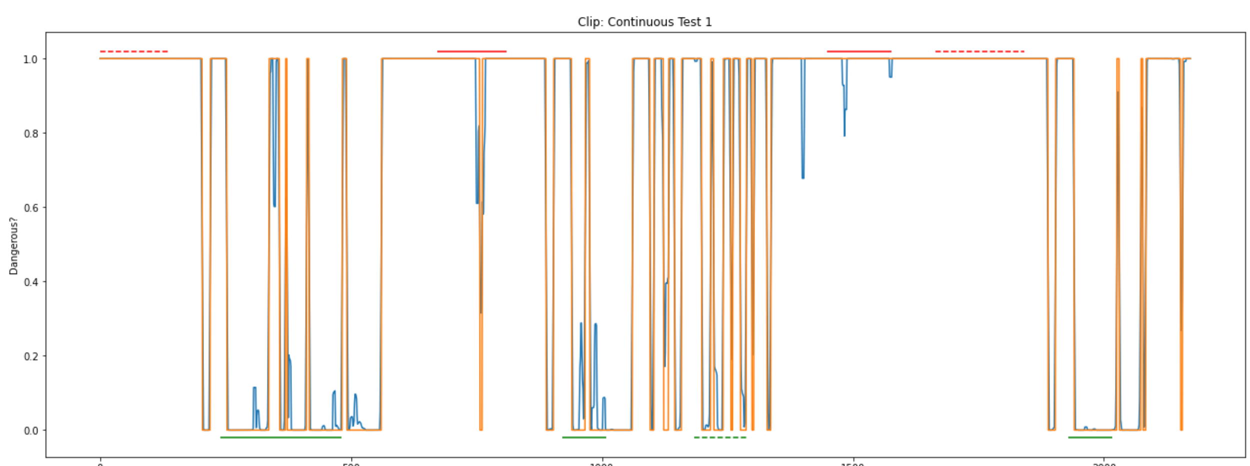

Figure 4: Qualitative evaluation of the clip detection network on the continuous recording

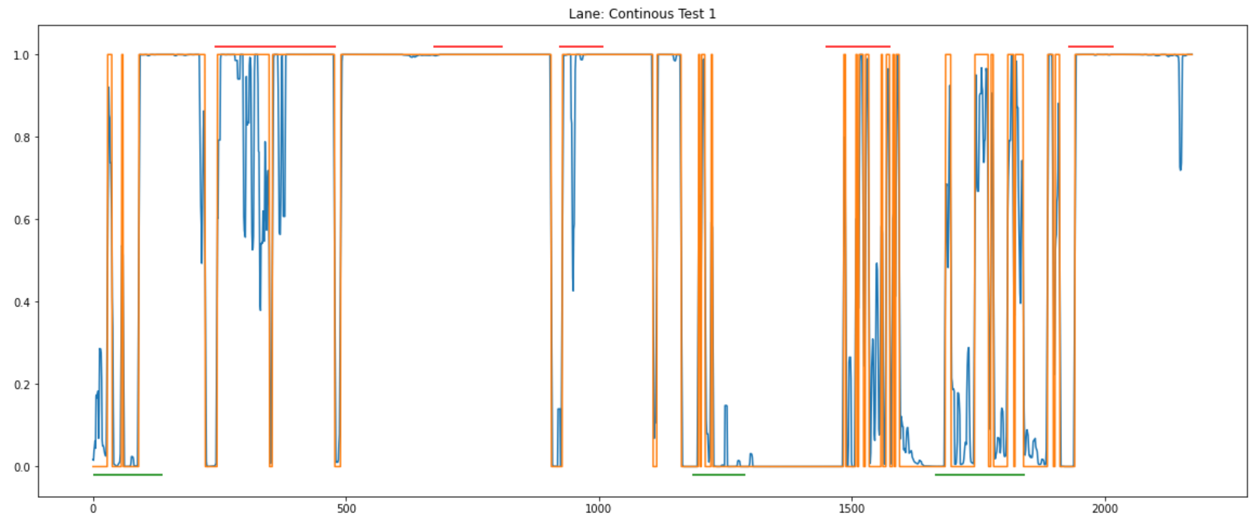

Figure 5: Qualitative evaluation of the lane detection network on the continous recording

7 Limitations

In this section, we will discuss the main limitations of the system. These are the limitations that we think are major and would like to solve in the future when the scope of the project expands.

The first and most crucial limitation of the project is that the dataset lacks on data points that contain background noise and music. As climbing gyms are extremely social environments with backgrounds music at all times, this is a highly unrealistic representation of the real-world use cases.

The second limitation of the project is that there was only one climber during the data collection process. Climbing is known to be a sport of high diversity in styles and skills. Having only a single climber in the datasets restricts the style and skill level of climbing that the algorithms can detect accurately.

The last major limitation we will discuss is the presence of multiple climbers at the same time. This will undoubtedly affect the performance of both clip and lane detection networks. In a real-world scenario, it's very likely that there are multiple climbers on the wall at the same time when the gym is busy.

We understand that there are many more limitations of this work, but these are the most imperative limitations as we see it.

8 Future Works

There are many aspects of the system that lack in order to be useful in real-world use cases. In this section, we will discuss possible future works to improve on these limitations that were discussed in the previous section.

8.1 More Extensive Dataset

The most straightforward next step is to have a much more extensive dataset. Ideally, this dataset would contain examples with background noise, music, climbers of different body types and skill levels, more auto belay lanes, and multiple climbers climbing at the same time.

We expect this dataset to be much harder to clean, but it will reflect the real-world in a much better.

8.2 Multiple Microphones

When there are many auto belay lanes, we will need to use multiple microphones for clip detection (one per each lane). The use of multiple microphones will significantly increase the accuracy of the lane detection network, which is bound to lose accuracy with more number of lanes.

We would also like to explore how many auto belay lanes can 3 microphones cover while also having a high accuracy in both of clip and lane detection networks. We suspect that the number is likely two to three folds from the number of microphones considering the auto belays are usually close together in one room.

8.3 Multiple Climbers

One of the more advanced features of the system would be to detect all lanes with climbers when there are multiple climbers on the wall at the same time. Although the current system design supports clip detection for multiple lanes at the same time (by having a separate network for each lane), it's restricted by the lane detection capacity of simultaneous multiple climbers.

To solve this issue, we would have to alter the design of the lane detection network. We could explore one network that can support multiple climbers or have a separate network for every lane that can detect if a climber is present on a certain lane.

9 Conclusion

Climbing is a fast-growing sport, and auto belays are an essential part of the climbing industry. In this paper, we presented BelaySound that can detect climbers who forgot to clip into their auto belay before climbing. Although the system shows a promising result in our tests, it still requires a lot of for any real-world use cases.

From our experiences in the climbing industry and our research, we found a great lack of machine-assisted technologies in the field. We hope this work is an inspiration for others to incorporate widely available computer science techniques to the climbing industry.

References

[1] Jagmohan Chauhan, Yining Hu, Suranga Seneviratne, Archan Misra, Aruna Seneviratne, and Youngki Lee. Breathprint: Breathing acoustics-based user authentication. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, pages 278–291, 2017.

[2] Climbing Business JournalClimbing Business Journal is an independent news outlet dedicated to covering the indoor climbing industry. Here you will find the latest coverage of climbing industry news. Gyms and trends 2021, Feb 2022.

[3] Xingzhe Song, Boyuan Yang, Ge Yang, Ruirong Chen, Erick Forno, Wei Chen, and Wei Gao. Spirosonic: monitoring human lung function via acoustic sensing on commodity smartphones. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, pages 1–14, 2020.