Body Pose Sonification for a View-Independent Auditory Aid to Blind Rock Climbers

Joseph Ramsay
University of Birmingham
joseph.f.ramsay@gmail.com

Hyung Jin Chang
University of Birmingham
h.j.chang@bham.ac.uk

Abstract

Rock climbing is a sport in which blind people have traditionally found it extremely difficult to excel due to the high degree of visual problem solving required, and also the requirement to climb with a sighted assistant. We present a system which automates the role of the sighted assistant in order to provide blind people with the freedom to climb and train on their own. We address climbing-specific limitations of a state-of-the-art skeleton tracking system, and discuss the ways in which we mitigated these limitations using post-processing techniques tuned specially for a climbing scenario. We also describe the auditory feedback system used to instruct the blind climber, and demonstrate that a user can learn to follow it in a relatively short time by showing a significant improvement in performance over just a few trials with the system.

1. Introduction

The motivation for this research is the growing popularity of indoor climbing in recent times [5]. To complete a route at a climbing gym, climbers must generally employ a great deal of visual problem solving to find the most efficient way to progress between foot- and hand- holds (called a 'move'). Sighted climbers enjoy the freedom to climb on their own, allowing for demanding training routines to be undertaken with very little logistical effort. Visually impaired climbers generally climb with a sighted assistant, whose role it is to direct the climber verbally from the ground. We investigate the feasibility of extending the same freedom to visually impaired climbers by replacing this assistant with a computational aid, with the lofty aim of reducing barriers to training, and hence helping to push standards for paraclimbers.

The sighted assistant performs many tasks which present a significant computational challenge. We chose to focus primarily on two particularly challenging aspects of the assistant's role:

- **Tracking the climber** - our system must be able to reason about the pose of the climber's body with respect to the wall in order to generate meaningful feedback. Previous attempts to characterise climbers' movements, such as ClimbAX [11] and a study by Sibella *et al.* [21], focus on large-scale quantities (such as a climber's centre of gravity) and as such cannot provide sufficiently detailed information about the positions of a climber's hands and feet. To our knowledge, vision-based skeleton tracking has not previously been applied with the specific purpose of aiding or augmenting the activity of climbing. We found that existing solutions for skeleton tracking did not perform well on the types of body positions commonly exhibited by rock climbers, with self-occlusions and crossed limbs being very common. This meant that extensive post-processing was required in order to track the climber's skeleton with sufficient accuracy to provide effective feedback.
- **Auditory feedback** - previous work aimed at assisting blind climbers has been primarily associated with the sensory substitution system Brainport [9], famously used by Erik Weihenmayer, the first blind person to climb Everest. We believe that the complexity of the output from such sensory substitution systems is not necessary for our specific use case, and that the amount of training required by such systems would present a very significant obstacle to many visually impaired climbers. Instead we chose to replicate the role of the sighted assistant more directly. The sighted assistant does not attempt to describe the entire scene to the climber, only those details pertinent to the next transition between holds. By following an approach similar to those used by more 'object-based' assistive systems such as those described in Section 2, we aimed to develop a system with a comparatively shallow learning curve which still delivers the information necessary for the climber to complete their route.

In this paper we outline our approach to tracking the climber, including post-processing techniques to address various limitations in skeleton tracking specific to this scenario, and a higher-level framework which allows for view-independent reasoning about a climber's progress through their route. We also present our method of delivering feedback to the user, which addresses a well-explored problem of directing a blind user to a target in space, with the added complication that it is crucial that they reach the target with the correct hand or foot. Finally we present a detailed evaluation of the effectiveness of this feedback based on user testing in a controlled environment.

2. Related Work

Object-based assistants: Our system can be thought of as a form of 'object-based' assistant. We use this term to refer to systems which help a blind person to locate an object in space. Object-based assistive systems rely on object recognition and tracking techniques to help direct users towards an object of interest. We are not so much interested in the object recognition techniques used in such systems (the targets which we are interested in are fixed in space, and can be obtained by the methods described in Section 4.1.3) as in the feedback systems used to indicate an object's location to a user.

One such system is presented by Schauerte *et al*. [17], which explores an approach to helping blind people find objects which they have previously lost. An important distinction between our system and Schauerte *et al*.'s is that, in our case, it is crucial that the object of interest is grasped with the correct hand or foot, something which their approach has no capacity to specify. WiYG, presented by Billah *et al*. [7], uses an approach very similar to ours to enable blind people to fill out printed forms. This high-precision haptic task is conceptually not dissimilar to the task of reaching to grasp a hold on a climbing wall. However WiYG operates within a much smaller area of space than our system, providing no large-scale instruction to direct the user to the general area of their target, and, like Schauerte *et al*.'s system, has no capability to specify which limb must be used to complete a task.

One important distinction when considering assistive systems for blind people is that of auditory versus tactile feedback. Two recent successful tactile systems are DLWV2 [18] and Palmsight [22]. Both use bulky hand-mounted devices which would be prohibitively restrictive if used while climbing. Mante and Weiland [14] evaluate the relative performance of tactile versus auditory feedback. Neither system showed any great advantage over the other, and as such we decided to use an auditory feedback system based on convenience.

The systems we have described up to now employ wearable devices (in fact, almost all such assistants do). Another important distinction for our work is that it is not appropriate for the system to be wearable. This is because it is entirely impractical for a climber to wear a camera or specialised device while climbing. The weight of a complex device would significantly add to the difficulty of climbing a route, and the risk of falling off and damaging the device makes the prospect of a wearable system less than ideal. Instead we chose to make use of a camera in a fixed location situated near the climbing wall. The fixed location of the camera meant that, once the system had been calibrated (as in functional requirement 4.1.3), the locations of all the holds in the image were easy to calculate by forward projection, meaning that object recognition techniques such as in the above systems were unnecessary.

Free-hands electronic travel aids: The system can also be considered a highly specialised form of electronic travel aid (ETA). Free-hands ETAs are ETAs in which feedback is delivered in such a way that the user is free to use their hands (an example of a non-free-hands system is a pair of gloves with vibrating elements in the fingertips [23]). Clearly we require a free-hands system for this research. A famous example of a free-hands ETA is *Navbelt* [19], proposed by Shoval *et al*. Navbelt delivers directional cues using a similar mechanism to our system however, like many free-hands ETAs [10] [12], it uses a wearable device. The nature of an ETA generally means that a wearable device is necessary in order to ensure constant monitoring of the subject. However even modern systems designed to operate in pre-determined environments, such as presented by Li *et al*. [13], use wearable devices, making them unsuitable for physically demanding tasks.

All ETAs studied, as well as the object-based systems, required a significant effort for the user to become comfortable using them. None of these systems aimed for immediate usability, but instead demonstrated their appropriateness by indicating an improvement in performance over time, either qualitatively or quantitatively (and often both). For this reason we aimed to demonstrate the same short-term learnability with our system, rather than attempting a longitudinal study to show that use of the system could become instinctive after prolonged use.

3. System Outline

The proposed system makes use of a device called a *Moonboard* (described in Section 3.1), which can be thought of as a subset of the sport of indoor bouldering. The basic use case is as follows: the user points a camera at the Moonboard and selects a problem from the Moonboard database. The user is then 'walked through' the selected problem, as the system tracks their motion through the route. They are informed of the position of the next hold whenever they complete a move, and given feedback as to the direction and proximity of the next hold as they move towards it. The system should detect when the user has fallen off the route, so that it can stop attempting to direct them (note that it is very common and safe to fall off when climbing on a Moonboard, as discussed in Section 3.3).

A number of climbing-related terms are defined here for the reader's understanding:

Hold	A small 3D shape made of plastic or wood which a climber can either grasp with their hand or push off with their foot
Move	A transition from one hold to another. In the system a move is considered to be the location of a hold in board space, and the joint (hand or foot) which needs to get there
Route	A sequence of moves, also referred to as a **problem**
Board space	Though not a term in most climbers' vocabulary, 'board space' is used to refer to the 2-dimensional grid on which the Moonboard holds lie. For example, the hold circled in red in the second image in Figure 1 is at position $(4, 0)$ in board space

3.1. The Moonboard

The Moonboard (shown in Figure 1), invented by Ben Moon in 2005, is a device used as a training tool by climbers worldwide. It consists of a single panel 2.44m high and 3.15m wide, with 142 holds arranged in an 18x11 grid. A 'problem' on a Moonboard is given by specifying a set of holds (e.g. $Start = F5, End = F18, Intermediates = \{E8, H10, F12, G13, D15\}$), where the indices are grid positions of the allowed holds. It is up to the climber to work out the transitions between holds, or *moves*. The totally standardised nature of the Moonboard makes it a good candidate for a task involving computer vision, as we already know where every hold is, so can fit a model of this to our image. Thousands of identical Moonboards exist worldwide, meaning the system can be used at many gyms (bouldering problems outside of the Moonboard have no 'standard'), as well as allowing climbers to follow detailed and structured training plans with very little logistical effort, wherever and whenever they like. A database of thousands of Moonboard routes exists online, meaning we can easily access many climbs for our user to try.

In images involving a climbing route, such as Figure 1, holds will often be indicated by coloured circles. The convention used by Moonboard is that holds circled in green are the starting holds (which the user must be holding before they begin to climb) and holds circled in red are the finishing holds, which the user must grasp with both hands to complete the route. Holds circled in blue are known as *intermediates* and can be used at any time.

3.2. Functional Requirements

A. Skeleton tracking - the system must be able to determine, from the camera stream, the position in image space of the climber's hands and feet. This is so that, when either a hand or a foot is moving towards a hold, effective feedback can be provided to indicate the proximity of the limb to the target hold.

B. Image-to-board projection - projecting the climber's body pose into board space allows us to provide instructions which are invariant to the position of the camera. ETAs are most intuitively followed when instructions are delivered with respect to the direction in which the recipient is facing. Since the climber can be assumed to be facing into the climbing wall, board space provides a good approximation to this.

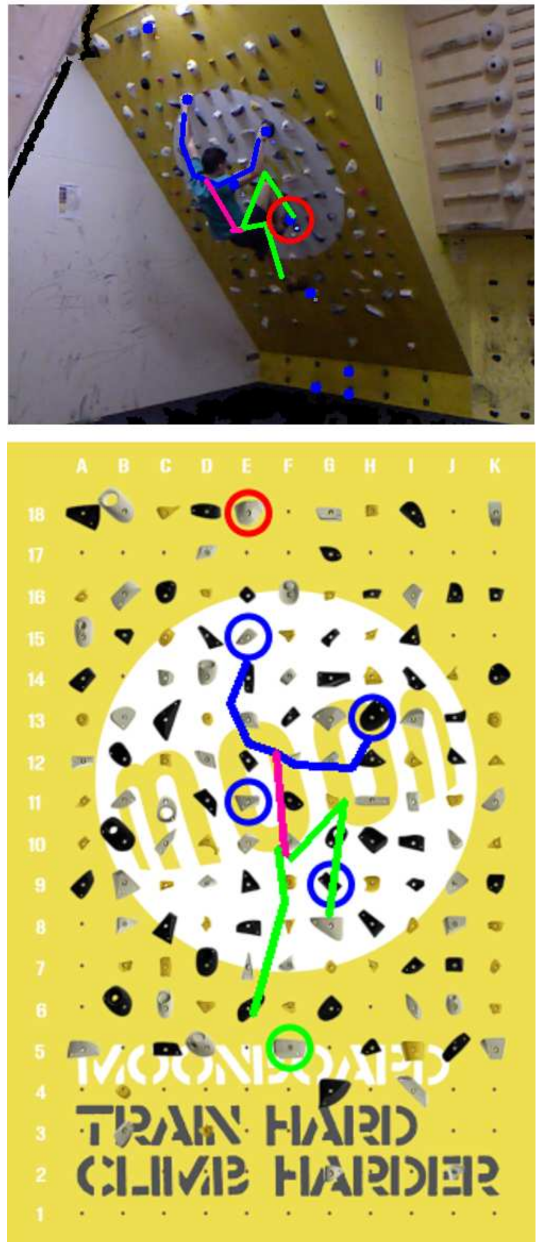

Figure 1. Image space view of a climber, and the corresponding projection to board space. The hold circled in red in the top image is the current goal

C. Fall detection & route completion - if a climber falls off a route, it is usual to try again from the start, not continue from where they fell off. The system should therefore be able to detect when the climber has fallen off the route, and stop giving feedback accordingly. Similarly, once the climber has held the finishing hold with both hands (the customary way to complete a boulder problem), the system should exit.

D. Auditory feedback

1. **Next hold instruction** - the direction and distance to the next hold, and the hand or foot to be used, should be announced to the climber on completion of a move.
2. **Proximity indication** - the distance and direction of the hold which the climber is currently moving to should be indicated continuously by some non-verbal signal.

E. Route Planning - in order for the system to be considered fully autonomous, it should include the functionality to predict the optimal sequence of moves required to complete a route. Despite being explored by Pfeil *et al*. [16] the general problem remains a very difficult task, and one with which even very experienced human climbers often struggle. There is however very often only one way of climbing a route on a Moonboard, and as such the climber who grasps a hold with their left hand which should have been grasped with their right will usually be unable to readjust and complete the route. This style of climbing means that the move sequence does not have be planned as the user climbs, but instead can be found out by a sighted climber, who will hand-annotate the routes in the system offline prior to running them.

3.3. Risk Analysis

Though the research involves indoor climbing, seen by some as an extreme sport, the risk to users of the system remains very low. The aim of the Moonboard is to provide an interesting training experience, not a long climb, and climbers' feet are generally no more than 1.5 metres from the ground, meaning any fall is likely to be short and easy to control. Boards are also surrounded by soft mats in order to cushion any falls, vastly reducing impact force in the event of a climber falling.

4. Methodology

The primary data pipeline which will be followed while a user is climbing is shown in Figure 2. This pipeline will become available once the user has calibrated the system (or loaded a pre-existing calibration) and selected a route. The first component, OpenPose, will respond to frames received from the camera, and pass the skeleton keypoints on to the pipeline. OpenPose keypoints tend to be noisy, so post-processing will be applied in the form of various filtering algorithms in order to smooth the keypoints, which will then be projected to board space. Board space coordinates will be used to update a model of the climbers progress in the chosen route, and the progress will be used to inform the feedback system on the type of sound to be emitted. In this section we describe our approach to each component of the pipeline.

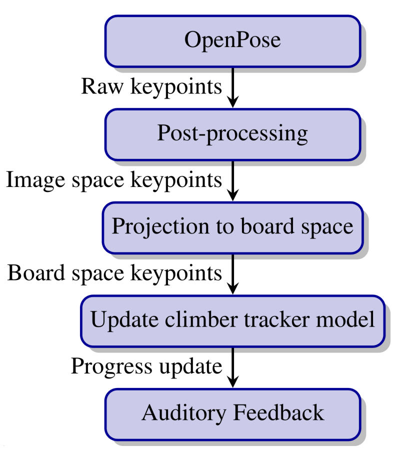

Figure 2. Data flow diagram for main system loop

A key component of the data pipeline is the climber tracker model, which will follow a goal-based loop. A goal is simply a move, as defined in Section 3. When the user selects a route, the goal is set to the first move in that route. When a new goal is set, it will be announced verbally to the climber. From that point until the goal is considered to have been achieved, a tone will be played indicating the proximity of the relevant joint to the target point in board space.

4.1. Tracking the Climber

4.1.1 Skeleton Tracking

We experimented with several different skeleton tracking technologies. Physical constraints on where the camera can be placed in relation to a Moonboard mean that the climber will always be facing away from the camera, and will generally be at least two metres from it. Due to this effect, and the unorthodox body positions commonly exhibited by climbers, we found that existing libraries such as OpenNI with NITE [1] were almost always unable to resolve the body pose of the climber. We found OpenPose [8] [2] to be the most effective solution, and so adopted this as the first component of the data pipeline. Keypoints were returned from OpenPose in 2D image space, which was then augmented with depth for reasons described in Section 4.1.3.

Two possible implementations of OpenPose were considered, CMU's original Caffe implementation [3] and a reimplementation using TensorFlow [2]. The TensorFlow version was chosen, as it can operate in real time at 30 frames per second, whereas the Caffe version is limited to around 10, and responsiveness of feedback was a critical priority for the project.

4.1.2 Post-Processing

Several post-processing filters were applied to the OpenPose keypoints in order to reduce noise. A Kalman filter, implemented in a similar to way to Shu *et al*. [20] and tuned specifically for the system, reduced small-scale noise. By applying a median filter to the velocity of keypoints, which provided the motion model in the Kalman filter, we introduced 'inertia' to keypoints, limiting the effect of large-scale spikes in acceleration characteristic of incorrect keypoint classification. Finally we made the observation that, due to the style of body positions found in climbing, in the case when a climber's hand or foot is occluded by some other part of their body, that joint is very often stationary. By running the above filters as normal in the case of an occluded joint, we would see the keypoint drift away from the climber, as its velocity from before it had been occluded would continue to be applied to it. In the case of an occluded joint, we instead just assume that the joint remains stationary, which contributes a surprisingly large improvement to

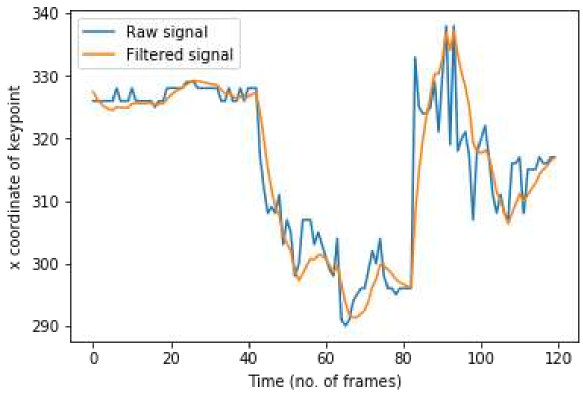

Figure 3. The effect of the combined filters. We show the x coordinate of the raw keypoint and smoothed keypoint for a climber's left hand from a four second video segment

accuracy. Figure 3 shows the effectiveness of this approach on a short segment of OpenPose output.

4.1.3 Projection to Board Space

In order to obtain a view-independent estimate of the climber's progress, we require a transformation T such that $\begin{bmatrix} x_b \\ y_b \end{bmatrix} = T \begin{bmatrix} x_i \\ y_i \end{bmatrix}$ where $\begin{bmatrix} x_b \\ y_b \end{bmatrix}$ is a point in board space and $\begin{bmatrix} x_i \\ y_i \end{bmatrix}$ is a point in image space. T may be non-linear, meaning that a two-dimensional image is not sufficient for the projection. Using a depth camera allows us to augment the image with depth information z_i, meaning we can express T as a perspective projection, where $\begin{bmatrix} x_i & y_i & z_i & 1 \end{bmatrix}^T$ are our *world* coordinates. Note that, while this approach seems slightly backwards (in conventional terms we are treating the 3D image as our *object*, and the board as our *image*), we found that it worked very well in practice, as shown in Figure 1. The user supplies the position in image space of a number of pre-determined calibration holds by clicking on the pixel containing them in a video feed from the camera. These are then augmented with the corresponding value in the depth map to give the *world space* coordinates. T is then calculated by minimisation using OpenCV [4], using the coordinates of the holds in board space as the *image space* coordinates.

4.1.4 Tracking Algorithm

The tracking model provides a high-level view of the climber's progress in their route. It is implemented as a simple goal-based loop, where a goal is the position of a hold in board space and the joint which must go to it. When a new goal is selected, it is announced verbally as described in Section 4.2.1. From that point until the goal is achieved, a tone is played as described in Section 4.2.2. If the climber falls off, or the final goal is achieved, the loop exits.

Fall Detection: In general the problem of fall detection is very complex [15] [6], however in climbing it is greatly simplified, as while the climber is progressing, they will be travelling upwards in board space (Moonboard routes never include any element of down-climbing). The origin in board space is in the top left, therefore once the climber's average velocity becomes positive in the y axis, we can conclude that they have fallen off, and stop tracking them.

Goal Completion: We use three thresholds to predict whether a goal has been achieved. T_{dist} is a threshold for the distance from the joint to the target hold, below which the joint is considered to be situated on the hold. T_v is a threshold for the joint's velocity, below which the joint is considered to to be stationary. A keypoint is considered to have achieved its goal if it is closer than T_{dist} to the target hold, and has been stationary (according to T_v) for T_{time} frames. We followed a trial-and-error procedure to tune the three thresholds. Precision is improved by reducing T_v and increasing T_{dist} and T_{time}, and recall is improved by doing the reverse. We found that $T_v = 6$ pixels per frame, $T_{time} = 5$ frames and $T_{dist} = 20$ pixels gave 100% recall, but around 70% precision when evaluated on our small set of sample videos.

4.2. Directing the Climber

4.2.1 Verbal Feedback

Verbal feedback gives a high-level instruction to the user as to the location of the next hold, and the hand or foot which needs to go to it. In order to provide some reference, the position of the hold is given in relation to the opposite joint to the one which is involved in the move. For example a typical verbal instruction is something like "A14 with your left hand. This is two right and horizontally in a line with your right hand". A14 refers to the grid position of the hold in Moonboard space. Using the fixed reference point (i.e. the stationary joint in the instruction) means that the climber can use their own proprioception to begin to find the hold, instead of relying solely on the grid position. The verbal feedback was generated by the Python `pyttxs` library.

4.2.2 Tonal Feedback

The verbal feedback provides a rough estimate for where in space the goal is, and the tone provides feedback by which the climber can make the small adjustments needed to reach the goal precisely. Two approaches for the tone were implemented. The first was a 'linear tone', whose pitch was linearly proportional to that of the distance from the relevant joint to the target hold. The frequency F is scaled between $F_{max} = 600$Hz and $F_{min} = 200$Hz, the range of frequencies deemed comfortable by our test users, by the following:

$$F = F_{max} - s * \left\| \begin{pmatrix} x_{target} \\ y_{target} \end{pmatrix} - \begin{pmatrix} x_{joint} \\ y_{joint} \end{pmatrix} \right\| * (F_{max} - F_{min}) \tag{1}$$

where s is a constant to bring the distance into the same order of magnitude as the frequency. Moves were generally no longer than 15 units in board space (the length of a move is measured by the distance to the target from the target of the last move which involved the joint associated with that move). Solving equation 1 for s with target-joint distance 15 gives $s \approx 11$. The second tone was a 'quadratic tone', which provided more fine-grained changes in pitch as the relevant joint moved closer to the target. The reason for this was that users found it easy to move the joint to the general

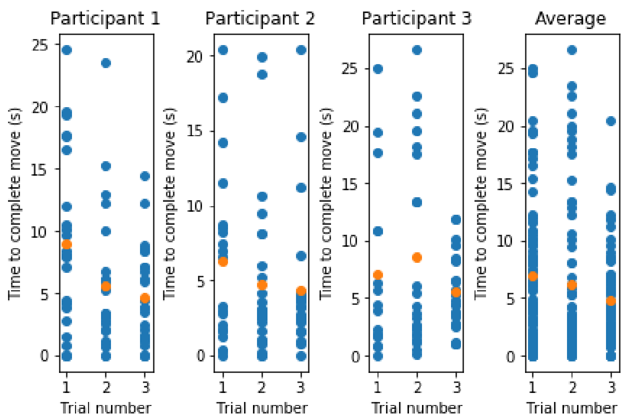

Figure 4. The times taken to complete a move in each trial, where a blue spot is the time taken for one move and orange spots are the means across that trial. We can see a clear improvement for participants 1 and 2, and an average improvement

vicinity of the target, but found the linear tone too imprecise to be able to make the small adjustments needed to complete the move. The quadratic tone aimed to address this issue:

$$F = F_{min} + s * \frac{1}{\left\| \begin{pmatrix} x_{target} \\ y_{target} \end{pmatrix} - \begin{pmatrix} x_{joint} \\ y_{joint} \end{pmatrix} \right\|^2} * (F_{max} - F_{min}) \tag{2}$$

Experimentally, we found $s = 0.1$ to provide good responsiveness. The 'angle' (the balance between left and right channels in a stereophonic system) of the tone was also controlled, in order to communicate which side of the joint the target is on, by $\theta = \frac{s*(x_{joint} - x_{target})}{2}$, where θ is the balance of amplitudes in the left and right channels of a stereophonic system ($\theta = 1$ for entirely left, 0 for right, and 0.5 for equal balance), and s is a scale factor to normalise $x_{joint} - x_{target}$ between -1 and 1. We found $s = \frac{1}{7}$ to be appropriate, as the horizontal length of a move rarely exceeded 7.

The tone was generated by the C++ `portaudio` library, which provides a cross-platform wrapper around the computer's sound card. We implemented a ROS node which constantly generates a sine wave, and responds to messages to control its frequency, angle and amplitude.

5. Experimental Results

5.1. Evaluation Method

For various logistical reasons we were unable to evaluate the system at a real climbing wall, instead simulating a climbing wall by directing the user to arbitrary spots on a vertical wall (see Section 5.5). The duration of the research did not allow for a long-term longitudinal study to investigate the transition from conscious to subconscious following of the auditory feedback. Instead we sought to demonstrate immediate learnability by showing that, over several sessions with the system, a user's performance in terms of the average time taken to achieve a goal showed significant improvement. As well as being a good indicator of the fluency with which a user is able to interpret the auditory feedback, the time taken to complete a move is pertinent to the case of a climber, as if the user cannot interpret and follow the feedback quickly and precisely, it is likely that they will just become tired and fall off the wall. We created several 'routes' in the simulated climbing wall designed to mimic the type of movements characteristic of climbing. We then tested the system on a sample of three participants over three trials each, recording the time to complete each move.

5.2. Participants

The participants were 2 females and one male between 20 and 23 years of age. One participant had around 5 years of climbing experience, and the other two were novice climbers. Interestingly the experienced climber showed no significantly greater aptitude for using the system than the two novices, suggesting that the greater degree of physical coordination commonly exhibited by experienced rock climbers is not necessary to become proficient using the system. We did not have an opportunity to evaluate the system with users who were totally unfamiliar with climbing, however our study is able to suggest that the system could offer an aid to experienced climbers who have lost their sight but would like to continue to climb.

5.3. Quantitative Evaluation

We can see from Figure 4 that the direction of the difference between means showed improvement between trial 1 (mean = 7.0s) and trial 3 (mean = 4.8s). The difference between means was statistically significant as determined by one-way ANOVA ($F(1, 1) = 6.8, p = 0.01$). The ANOVA shows that the improvement was statistically significant ($p < 0.01$). We conclude that it is reasonable to expect the performance of a user of our system to improve with experience, however the evidence available to us is not sufficient to show that use of the system becomes truly intuitive.

5.4. User Study

Participants were asked to complete a short questionnaire following their final trial. The results are summarised below:

Question	**Mean Answer**
In your first trial, how easy did you find the system to use? (1 =very difficult, 5 = very easy)	2
In your final trial, how easy did you find the system to use? (1 = very difficult, 5 = very easy)	5
In your first trial, how intuitive did you find the system to use? (1 = very non-intuitive, 5 = very intuitive)	3.33
In your final trial, how intuitive did you find the system to use? (1 = very non-intuitive, 5 = very intuitive)	3.67
How important did you find the spoken instructions? (1 = unnecessary, 5 = crucial)	4.33
How easy did you find it to interpret the spoken instructions? (1 = very difficult, 5 = very easy)	3
How important did you find the tone? (1 = unnecessary, 5 = crucial)	4.67
How easy did you find it to interpret the tone? (1 = very difficult, 5 = very easy)	4.33

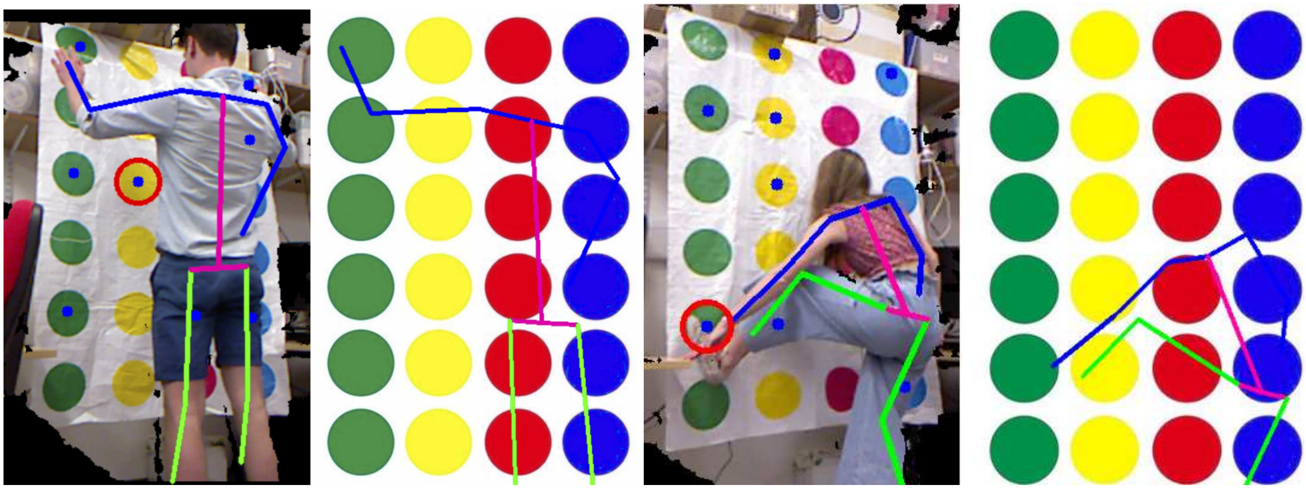

Figure 5. Image space view of a Twister player, and the corresponding projection to Twister space. As in Figure 1, the hold circled in red in the left image is the current goal. Note that the black patches represent blind spots in the depth camera.

1. The users reported that the system became easier to follow as they became more familiar with it, which explains the improvement in the time taken to complete each move. They did not find that it became more "intuitive", though we did not expect to see this effect over such a short trial.
2. Participants occasionally found the spoken feedback confusing. We describe in Section 4.2.1 how target positions are given with respect to the position of another hold which the opposite limb is supposedly currently holding. This was necessarily the case when the user was climbing, for example if a move involved a left hand, we could safely assume that the right hand would not have moved since it was last involved in a move, otherwise the climber would have fallen off. However the same assumption could not be made about users in our simulated climbing wall, as there was no physical requirement to keep the hand or foot in position once the move was over. The limb would often drift from its original target, meaning that feedback with respect to this point was sometimes misleading. It is expected that the reference position would be more useful when the user of the system was a climber, however we would have to conduct a more thorough study to establish this.

5.4.1 Feedback Models

The above analysis is for the quadratic tone described in Section 4.2.2. Though we initially experimented with the linear tone, all participants stated very early on that they found it extremely difficult to make small adjustments in order to achieve the target when following the linear tone, and that the quadratic tone was much easier to follow. For this reason, we adopted the quadratic tone for the remainder of the study, and abandoned the linear tone.

5.4.2 Self-Occlusion Avoidance

One interesting observation that we made during the evaluation process was that, as they became more familiar with the system, users became more adept at dealing with errors in skeleton tracking. In cases where a joint could not be found in the image, possibly due to self-occlusion, the joint would simply remain stationary in the position in which it was last detected. This manifested to the user as a constant tone, regardless of how much they moved the limb being tracked. The participants all learned to recognise this symptom, and were observed to experiment with different body positions in order to allow the lost joint to be found again.

5.5. Applying to Twister

We also experimented with applying our system to another whole-body reaching task, namely the popular game *Twister*. We were able to adapt the projection model to Twister very easily, providing as calibration points the centres of each spot on the Twister mat, and performing the calibration in the same way as for the Moonboard. As well as showing that our system is robust enough to handle a variety of complex reaching tasks, this extension provided the 'simulated climbing wall' described in Section 5.1. We achieved this by hanging the Twister mat on a wall, and creating routes to mimic the type of body positions commonly exhibited by climbers. Figure 5 shows the system being used in this way.

6. Conclusion

The presented system has been shown to be an effective approach to assisting blind users in whole-body reaching tasks. The view-independent model of the climber's progress provides a robust mechanism to reason about body pose in the complex domain of rock climbing. By showing a statistically significant improvement in user performance over a small period of time we have shown that users can become familiar with the system with relatively little training. Though not evaluated at a real climbing wall, we believe that we simulated a climbing wall closely enough to present our system as a viable proof-of-concept in a climbing environment. We would be extremely interested to investigate the system's effectiveness at a real climbing wall over a much longer period of time and with a larger number of participants to determine whether it is possible for users to begin to use it truly intuitively.

Another interesting extension would be to augment the sensor model of the Kalman filter with data from wearable devices such as accelerometers on the wrists and ankles of the climber. We believe this would be an effective way to mitigate the issue of self-occlusion of the climbers joints. Instead of assuming that occluded joints remain stationary, which worked reasonably well but did cause issues during user testing, acceleration information could be used to further inform the measurement update step in the case that positional information was not available.

Acknowledgement: This work was supported by UoB-Wellcome Trust Institutional Strategic Support Fund: Digital Health Pilot Grant.

References

[1] https://structure.io/openni.

[2] https://github.com/ildoonet/tf-pose-estimation.

[3] https://github.com/CMU-Perceptual-Computing-Lab/openpose.

[4] https://docs.opencv.org/2.4/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html. Accessed: 31 March 2019.

[5] N. Berry. Social climbers - the evolving indoor climbing industry. https://www.ukclimbing.com/articles/features/social_climbers_-_the_evolving_indoor_climbing_industry-10953. Accessed: 22 October 2018.

[6] Z.-P. Bian, L.-P. Chau, and N. Magnenat-Thalmann. A depth video approach for fall detection based on human joints height and falling velocity. *International Coference on Computer Animation and Social Agents*, 2012.

[7] S. M. Billah, V. Ashok, and I. Ramakrishnan. Write-it-yourself with the aid of smartwatches: A wizard-of-oz experiment with blind people. In *23rd International Conference on Intelligent User Interfaces*, pages 427–431. ACM, 2018.

[8] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7291–7299, 2017.

[9] Y. Danilov and M. Tyler. Brainport: An alternative input to the brain. *Journal of integrative neuroscience*, 4:537–50, 01 2006.

[10] D. Kalampalikis and K. Moustakas. Design of a vision substitution vibrotactile vest for the visually impaired. In *Proceedings of the 10th Hellenic Conference on Artificial Intelligence*, page 52. ACM, 2018.

[11] C. Ladha, N. Hammerla, P. Olivier, and T. Ploetz. Climbax: Skill assessment for climbing enthusiasts. 09 2013.

[12] A. Lakshmanan, S. G. Devi, M. M. Nisha, and M. Dhanalakshmi. Outdoor obstacle detection module to assist visually challenged. In *Advances in Communication, Devices and Networking*, pages 473–481. Springer, 2019.

[13] B. Li, J. P. Munoz, X. Rong, Q. Chen, J. Xiao, Y. Tian, A. Arditi, and M. Yousuf. Vision-based mobile indoor assistive navigation aid for blind people. *IEEE transactions on mobile computing*, 18(3):702–714, 2019.

[14] N. Mante and J. Weiland. Visually impaired users can locate and grasp objects under the guidance of computer vision and non-visual feedback. volume 2018, pages 1–4, 07 2018.

[15] Y. Nizam, M. N. H. Mohd, and M. M. A. Jamil. Human fall detection from depth images using position and velocity of subject. *Procedia Computer Science*, 105:131 – 137, 2017. 2016 IEEE International Symposium on Robotics and Intelligent Sensors, IRIS 2016, 17-20 December 2016, Tokyo, Japan.

[16] J. Pfeil, J. Mitani, and T. Igarashi. Interactive climbing route design using a simulated virtual climber. *SIGGRAPH Asia 2011 Sketches, SA'11*, 12 2011.

[17] B. Schauerte, M. Martinez, A. Constantinescu, and R. Stiefelhagen. An assistive vision system for the blind that helps find lost things. In K. Miesenberger, A. Karshmer, P. Penaz, and W. Zagler, editors, *Computers Helping People with Special Needs*, pages 566–572, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[18] M.-L. Shih, Y.-C. Chen, C.-Y. Tung, C. Sun, C.-J. Cheng, L. Chan, S. Varadarajan, and M. Sun. Dlwv2: A deep learning-based wearable vision-system with vibrotactile-feedback for visually impaired people to reach objects. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1–9. IEEE, 2018.

[19] S. Shoval, J. Borenstein, and Y. Koren. The navbelt-a computerized travel aid for the blind based on mobile robotics technology. *IEEE Transactions on Biomedical Engineering*, 45(11):1376–1386, 1998.

[20] J. Shu, F. Hamano, and J. Angus. Application of extended kalman filter for improving the accuracy and smoothness of kinect skeleton-joint estimates. *Journal of Engineering Mathematics*, 88(1):161–175, 2014.

[21] F. Sibella, I. Frosio, F. Schena, and N. Borghese. 3d analysis of the body center of mass in rock climbing. *Human Movement Science*, 26(6):841 – 852, 2007.

[22] Z. Yu, S. Horvath, A. Delazio, J. Wang, R. Almasi, R. Klatzky, J. Galeotti, and G. Stetten. Palmsight: an assistive technology helping the blind to locate and grasp objects. *Carnegie Mellon University, Pittsburgh, PA, Tech. Rep.*, pages 16–59, 2016.

[23] J. Zelek, R. Audette, J. Balthazaar, and C. Dunk. A stereo-vision system for the visually impaired. *University of Guelph*, 1999, 1999.