Use of Markerless Pose Estimation for Clinical Gait Analysis of Prosthetic Users

Anthony Cimorelli

Shirley Ryan AbilityLab Prosthetic Residency Research Report

Acknowledgements:

Thank you to the following individuals for their help with this project: R. James Cotton, Emoonah McClerkin, Laura Miller, Andrea Ikeda, Suzanne Finucane, Ann Simon, Shenan Hoppe-Ludwig, Kristi Turner, Robert Lipschutz, Shirley Ryan AbilityLab Outpatient Physical Therapists, Center for Bionic Medicine

Funding Support:

The contents of this report were developed under grants from the Research Accelerator Program of the Shirley Ryan AbilityLab and the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR grant number 90REGE0003). NIDILRR is a Center within the Administration for Community Living (ACL), Department of Health and Human Services (HHS). The contents of this report do not necessarily represent the policy of NIDILRR, ACL, or HHS, and you should not assume endorsement by the Federal Government.

Abstract

Introduction: Several gait analysis techniques have been used to study prosthetic gait for many years. Previous studies have been limited by small sample sizes and restricted to laboratory settings. Recent advances in computer vision have led to significant progress in the ability to analyze video and extract human movements, an approach called markerless pose estimation. The use of markerless pose estimation techniques on prosthetic users has not been widely studied. The goal of this study was to determine the possibility and identify challenges of using markerless pose estimation techniques in conjunction with a custom gait transformer to perform gait analysis in prosthetic users in a clinical setting.

Subjects: 21 individuals with lower limb amputations were recruited and provided written informed consent. Level of amputation and prosthetic componentry varied between subjects.

Procedures: Data was collected during normal prosthetic and/or physical therapy visits at Shirley Ryan AbilityLab. Custom wearable sensors were placed on shank and thigh of the prosthetic limb. Video and sensor data was obtained through a custom app on an Android cell phone as mobility activities occurred naturally throughout the clinical visit. Black tape was placed on the floor at 10-meter distances to calculate ground truth walking velocity.

Data Processing: Processing of the videos was performed using a custom gait analysis pipeline. The pipeline has three main steps: subject identification, 2D keypoint detection, and gait transformer outputs. The gait transformer has been trained to output gait event timing and spatio-temporal gait parameters. Ground truth cadence was measured from the wearable sensors.

Results: Subject identification and tracking worked well with 263/270 (97%) of videos identifying the person of interest for >90% of the video. Several factors were identified that appear to be affecting 2D keypoint detection on prosthetic users including: whether the prosthesis was visible or covered by clothing, level of amputation and the type of prosthetic socket. The gait analysis system's ability to detect walking was limited, and walking was not detected at all in two individuals. Walking velocity (R^2= 0.740, r=0.894) and cadence (R^2= 0.694, r=0.843) were measured from the video with relative accuracy. The feasibility of detecting longitudinal gait changes was also demonstrated.

Discussion/Conclusion: We showed that it is possible to measure quantitative gait data from video recorded on a cell phone camera in a clinical setting. Differences in limb and gait characteristics between prosthetic users and the individuals that the pre-trained algorithms and the gait transformer were trained on appear to be affecting results. Best results were obtained for individuals who's prostheses appear more like those with intact limbs. Additional work is needed to improve the system's ability to work properly on prosthetic users. However, with improvements in the ability to locate 2D joints and accurately detect walking, this system shows the potential to be used quickly and easily in a clinical setting with minimal equipment or training required. Future work will attempt to measure additional spatial-temporal and kinematic gait parameters.

Introduction

It is estimated that there are currently two million people living with limb loss in the United States[1]. That number continues to increase and it is projected that approximately 185,000 individuals in the United States undergo an amputation of an extremity each year[2]. With the increase in amputations each year, more individuals are seeking prosthetic care.

For individuals with a lower limb amputation, various prosthetic components are prescribed and the resulting gait pattern is very complex with many factors affecting how an individual ambulates using a prosthesis [3,4]. It is well known that prosthetic users ambulate differently than able bodied individuals [3]. As a result, for many years, researchers have been attempting to better understand differences in gait characteristics of prosthetic users with different levels of lower-limb amputation [5]. As the number of people using prostheses continues to increase and prosthetic technology advances, it is important for us to develop methods to study the way that prosthetic users are moving and interacting with their environment.

Many gait analysis systems exist and are used regularly by researchers and clinicians. The gold standard is a laboratory with optical motion tracking and force plates [6]. While these systems produce precise results, they are expensive and time consuming to use, require the participant to wear many markers or sensors and necessitate extensive training to be able to operate and process the data they produce. Other systems including plantar pressure-based measurement techniques or systems that use wearable sensors that are easier to set up but typically produce less precise results and often require extensive calibration. Other devices, such as the iPecs load cell, have been developed and validated specifically for use in the prosthetic population[7,8]. Most of these devices use some variation of a load cell that is mounted distal to the prosthetic socket. While these designs produce useful data, they are expensive and require the prosthetist to install the componentry on the individual's prosthesis. Due to the many limitations associated with different gait analysis systems, prosthetists typically rely on observational gait analysis in the clinic when performing dynamic alignment adjustments to a patient's prosthesis [4].

Prosthetic gait has been studied in depth for many years [9–12]. Previous studies have used a variety of gait analysis systems to compare different prosthetic components, mass distribution, alignment of components or the effects of walking surface. Most of these studies have been performed in a controlled laboratory setting using the traditional gait analysis systems mentioned above. While these studies have assisted in understanding prosthetic gait, they have been limited by small sample sizes and the use in a laboratory setting.

Clinically, performance-based outcome measures are used to track a patient's progress over time [13]. Previous studies have shown performance-based outcomes to be reliable and valid for use in individuals with lower limb amputations [14]. However, these outcomes are not always used routinely in clinical practice as they are often time consuming and challenging to perform. They usually require the participant to complete some sort of timed walking test or a variety of different balance and mobility activities. Furthermore, performance-based outcome measures typically measure constructs such as walking ability, endurance and functional mobility. However, they cannot quantitatively measure values to determine how one's quality of gait is improving beyond walking speed. The ability to easily measure gait parameters such as spatio-temporal measures and joint kinematics would allow quantitative measurement to show how one's quality of gait is changing over time.

Recent advances in machine learning applied to computer vision have led to significant progress in the ability to analyze video and extract human movements, an approach called markerless pose estimation [15]. These pose estimation algorithms follow a taxonomy of approaches ranging from locating joint positions in two-dimensional images to attempting to infer the kinematic joint trajectories from videos[16]. The ability to calculate kinematic data from easily obtainable video would greatly facilitate human movement science and rehabilitation. More specifically, this process would enable a wider range of objective outcome measures that could be used to track performance in the therapy gym, clinic, in the community or at home. Recent work has demonstrated these approaches can measure clinically relevant outcomes such as walking speed, cadence, and knee angles in individuals with cerebral palsy [17]. As well as determine clinical rating scales in individuals with Parkinson's Disease [18]. In addition, previous work has shown that it is possible to measure joint kinematics when using multiple cameras and wearable sensors in a home or clinical setting [19,20].

Key to the success of these machine learning, sometimes called deep learning, algorithms is that they are trained on large amounts of data. While there are many existing approaches and datasets to train these algorithms, they all share a common limitation: that they are trained and evaluated on datasets of able-bodied individuals. Prosthetic limbs can have a range of appearances depending on the individual's remaining anatomy and specific prosthetic componentry used. Typically, prosthetic devices do not appear, nor necessarily move, like an intact limb, however sometimes a cosmetic cover will be incorporated to give the prosthesis a more anatomical appearance. In addition, clinically it is well known that prosthetic users ambulate differently than able-bodied individuals [3]. In machine learning, algorithms can fail to generalize to data that is different from the training data. As the use of markerless pose estimation techniques on prosthetic users has not been widely studied, it is unclear whether the differences in limb characteristics between prosthetic users and able-bodied individuals will lead to poor generalization with currently established pose estimation algorithms.

Therefore, the goal of this study was to determine the possibility and identify challenges of using currently available markerless pose estimation techniques to perform gait analysis on prosthetic users in a clinical setting. Prosthetic users were tracked and recorded during their regular prosthetic and/or physical therapy clinical visits to determine the performance of this type of gait analysis system in a non-laboratory setting. Some participants were filmed over the course of several visits to determine the ability for use as an outcome measure for longitudinal assessment. Custom wearable sensors were used to assess the accuracy of the video-based gait characterization. Based on the outcomes of this evaluation, the contributions of this study are that we identified factors that affect the use of pose estimation techniques on prosthetic users, determined the accuracy of measuring walking velocity and cadence from video and demonstrated the potential to use this system as an outcome measure to track longitudinal changes.

Methods

Participants

A convenience sample of twenty-five participants with lower limb amputations were recruited. Participants were included if they were between the ages of 18-95, had a unilateral or bilateral lower limb amputation, were classified as a K2-K4 functional level of ambulation with a prosthesis, were being seen for prosthetic care and/or outpatient physical therapy at Shirley Ryan AbilityLab, and were English speaking. Participants were excluded from the study if they had a significant new injury that would prevent the use of a prosthesis, cognitive impairments sufficient to adversely affect understanding of or compliance with study requirements or ability to give informed consent, or any other significant comorbidity that would prevent them from using their prostheses or prevent acquisition of useable data by researchers. This study was approved by the Northwestern University Institutional Review Board (IRB) (STU00215263). All individuals provided written consent prior to participation in the study.

Data Collection

Data was collected while participants completed any mobility activity during their prosthetic clinic visits or therapy visits at Shirley Ryan AbilityLab. Prior to any data collection, black tape was placed on the floor in the clinical areas at ten-meter distances to allow retrospective annotation of ten-meter walk times. The wearable sensors, a custom 9-axis inertial measurement unit developed at Shirley Ryan AbilityLab that acquire acceleration, gyroscope, and magnetometer data, were calibrated prior to each session as previously described [21]. Sensors were applied on the participant's prosthetic side at the beginning of each appointment. One sensor was placed on the lateral side of the shank as close to the proximal attachment of the prosthetic foot as possible, and one sensor was placed centered on the lateral side of the thigh (Image 1). No sensors were placed on the intact limb. Therefore, for unilateral participants two sensors were used and for bilateral participants four sensors were used. Sensors were applied directly to the prostheses with adhesive Velcro. For transtibial participants the thigh sensor was attached with an elastic strap. Once sensors were applied at the beginning of the appointment, they remained on until the end of the appointment to minimize any interference with routine clinical care.

Video data was then obtained through a custom app on an Android cell phone camera in both planes (sagittal, frontal) if possible as mobility activities occurred naturally throughout the clinical visit. The researcher collecting video data would ambulate a few feet behind, in front of or next to the participant to collect video data in both planes. A gimbal (DJI Osmo Mobile 3) was used to stabilize the cell phone camera (Image 1). Sensor data was acquired by the same app. Any prosthetic adjustments made during the prosthetic visits were documented.

The number of clinical appointments in which data were collected varied between participants based on their clinical schedule. Data were collected on participants one to ten times. For people that were seen multiple times during physical therapy, the target frequency of data collection was one time per week to capture longitudinal changes in gait. Apart from placing the sensors, and capturing video and sensor data, all the study procedures represented the standard clinical treatment of individuals with a prosthesis.

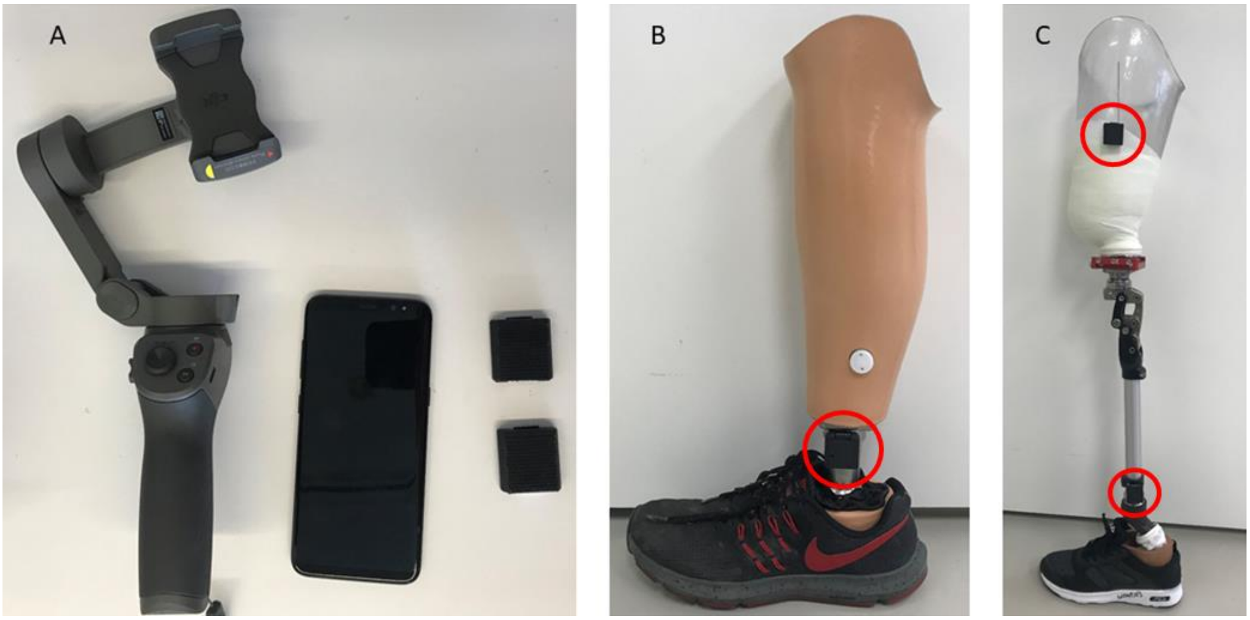

Image 1: Gait analysis system hardware. A: Android cell phone, two wearable sensors and gimbal. B: One sensor placed on the shank of a definitive transtibial prosthesis with adhesive Velcro. C: Two sensors placed on the shank/thigh of a diagnostic transfemoral prosthesis.

Data Processing

Post data processing of the videos was performed using a custom gait analysis pipeline that has previously been described [22]. Figure 10 in the appendix shows an overview of the gait analysis pipeline. The pipeline can be broken up into three main steps: subject identification, 2D keypoint detection, and gait transformer outputs.

Step 1: Subject Identification: The initial stage of the gait analysis pipeline is for the system to identify all individuals in the frame and surround them with a bounding box. In most videos an algorithm called DeepSort was used [23]. However, in a small subset where this method did not track the individuals properly, TraDeS [24] was used instead. Next, the research participant was manually identified by the author as the person of interest to track for the duration of the video.

Step 2: 2D Keypoint Detection: Once the person of interest was labeled, anatomical joints were identified as 2D keypoints in each frame of the videos using a top-down method. We used the MMPose library for this [25], which provides many pretrained algorithms, and specifically we used an HRNet architecture [26] with a distribution-aware coordinate representation [27] trained on the Microsoft Common Objects in Context dataset [28]. The software generates a confidence level of the location of each joint on a scale of 0-1, with higher values indicating that the algorithm was more confident in locating the joint. These values are averaged for the duration of the video when the person of interest is identified in the frame. While the software generates a confidence level for each joint, Ankle Keypoint Confidence and Knee Keypoint Confidence will be discussed in the results of this paper as the individuals included were utilizing lower extremity prostheses. The 2D keypoints are then lifted to 3D using GastNET [29].

Step 3: Gait Transformer Outputs: Once the 2D keypoints had been lifted to 3D, the videos were processed through a custom gait transformer that has previously been trained to analyze the videos frame by frame and measure several different gait parameters such as walking velocity, cadence, step length and various joint kinematics. The gait transformer can detect when an individual is ambulating ("Detecting Walking") and averages the gait measures over the time that it identifies the individual of interest is ambulating. The criteria for walking detection consists of several factors including that the bounding box for the individual is not at the edge of the frame, as this can cause errors in the tracking. It also assigns a lower probability if the 2D keypoint detection reports a lower confidence for either ankle keypoint. As described below, we process the gait transformer timing outputs with an Extended Kalman Filter, and the probability that walking is detected is reduced when the error between the predicted timing outputs and the measured timing outputs are higher. The probability that walking is detected also gradually begins to decrease as the cadence drops below 30 steps per minute or if there is a wide variability in the cadence. All of these criteria are designed to make our detection of walking conservative, and to not report results when the tracking, keypoint detection, or gait transformer outputs may be inaccurate. Refining these criteria to balance this is an area of future research. The detected walking measure is reported as the percent of the video ("Fraction Frames") that the system detected the individual was walking. While the gait transformer can output several different gait parameters the following measures will be discussed in the results of this paper: Velocity ("Gait Transformer Velocity"), Cadence ("Gait Transformer Cadence") and Kalman Error. Kalman Error is the error calculated once an Extended Kalman Filter was applied to the gait transformer output [30]. It is an "Ad-hoc" measure used to analyze the quality of the gait transformer output on a continuous scale from zero to one. A lower Kalman error value indicates a higher quality gait transformer output while a higher Kalman error value indicates a lower quality gait transformer output.

Further information about the details of the gait analysis pipeline is beyond the scope of this study. For more information on the data processing methodology, please reference [22].

Video Annotation

After the videos were run through the initial stages of the gait analysis pipeline and 2D keypoints were identified by the algorithms, the videos with 2D keypoints superimposed were assessed by the author and the following categories were manually annotated.

Annotation Category	Explanation and options for each category
Activity	Videos were recorded while participants were completing any mobility activity during their clinical appointment. Activities included level walking, treadmill walking, stairs, floor recovery and various balance and agility activities.
Assistive Device	Use of an assistive device by the participant during mobility activities varied. Devices included a walker, cane (single point and quad), axillary crutches, forearm crutches and a harness.
Prosthesis Visible	Whether or not the participant's prosthesis was visible or covered by clothing varied. Yes indicated that the participant's pant legs were rolled up so that a portion of the prosthetic limb was visible. No indicated that the participant's pant legs were down so that no portion of the prosthetic limb was visible in the video.
View	The plane of view the video was captured in varied. View included Frontal, Sagittal, or Both. Both indicates that the video was taken from an angle between planes or plane of view changed during the video.
Prosthetic Socket	Type of prosthetic socket utilized by the participant. Diagnostic: clear plastic with fiberglass wrap reinforcement. Definitive: laminated material with matched skin tone, carbon fiber or other custom appearance.
Prosthetic Keypoint Quality	To evaluate how well the pre-trained pose estimation algorithms work on prosthetic limbs, each video was manually assessed for 2D keypoint detection by the author using clinical judgement based on the following scale: 3: Keypoints correctly locate the joint and track well throughout the video 2: Keypoints correctly locate the joint but tracking is intermittently inaccurate throughout the video 1: Keypoints do not correctly locate the joint and tracking is frequently inaccurate throughout the video See Image 2 below for example of ratings.
Annotated Velocity	To evaluate the accuracy of the gait transformers velocity measure, average velocity was calculated by manually recording the time it took participants to ambulate 10 meters. Two black lines were placed in the clinical areas at 10-meter distances apart. The author paused the videos and recorded the exact time the individual crossed each line to determine average velocity over that distance. Note this is a different calculation of velocity than the standard 10-meter walk test used clinically as an outcome measure.

Image 2: Example of Prosthetic Keypoint Quality Ratings. A: Prosthetic Keypoint Quality 1 (Sagittal Plane, Prosthesis Visible). Note that the left ankle keypoint is identified on the right ankle and the left knee keypoint is identified mid-thigh. B: Prosthetic Keypoint Quality 2 (Frontal Plane, Prosthesis Visible). Note that the right ankle and knee keypoints are identified too proximal. C: Prosthetic Keypoint Quality 3 (Frontal Plane, Prosthesis Visible). D: Prosthetic Keypoint quality 3 (Frontal Plane, Prosthesis Covered).

Sensor Data Processing

The gyroscope data from the sensor mounted on the ankle was used to detect swing phase events and validate the cadence measurements. Specifically, we took a gyro channel corresponding to rotation in the sagittal plane (the z-axis with our position) and flipped the sign when it was on the right, so a positive rotation corresponded to the shank swinging forward (or equivalently rotating with the top of the shank moving backward relative to the ankle). We low pass filtered this and then took the detected positive peaks in the waveform that were above one-quarter of the 99% percentile of the gyroscope value. Visual inspection confirmed this heuristic worked reliably across many participants. Note that while this detects a swing phase event, the exact time that swing phase starts and stops is not calibrated to align with the toe off and heel contact events.

Statistical Analysis

Statistical analysis was performed in Python using the statsmodel package and are described below.

Results

Subject Demographics

For the results of this study, only videos of participants walking on a level surface were analyzed and other activities acquired such as treadmill walking, floor recovery and agility drills were not included. Four of the initial twenty-five participants were excluded as data was not collected of them walking on a level surface: two participants only had data of other activities and two participants were consented, but future clinical visits did not overlap with study timing and therefore no ambulation data was collected. Therefore, data from twenty-one participants was analyzed and described in this paper. Table 1 shows an overview of the demographic and prosthetic information for all the subjects included. Of the twenty-one subjects: sixteen were male and five were female, eleven were individuals with a unilateral transtibial amputation (TT), two were individuals with bilateral transtibial amputations (B-TT), four were individuals with a unilateral transfemoral amputation (TF), two were individuals with a unilateral knee disarticulation amputation (KD), one was an individual with bilateral transfemoral amputations (B-TF) and one was an individual with a unilateral hip disarticulation amputation (HD). In the results below, the individuals with knee disarticulation amputations were included in the transfemoral group. Etiology of amputation varied between subjects. Seventeen subjects were classified by the referring physician as MFCL K3 level ambulators and three subjects were classified as K2 level ambulators. One subject did not have a K-level reported. The average age was 55.7 years with a range from 22 – 77 years. The average height was 174.5cm with a range from 152 – 190cm. The average weight was 84.5kg with a range from 55 – 120kg. Prosthetic suspension, interface, foot and knee varied between participants. Five subjects utilized a test or diagnostic prosthetic socket, twelve subjects utilized a definitive prosthetic socket, and four subjects began data collection utilizing a diagnostic socket but were transitioned to a definitive socket during the data collection period so data were included for both sockets.

Table 1: Subject Demographics and Prosthetic Information.

Subject ID	Age	Gender	Height (cm)	Weight (kgs)	Level	Side	Etiology	K-Level	Socket	Suspension	Interface	Knee	Foot
101	50	Male	183	101	TT	Left	Trauma	3	Test	Sleeve	Gel liner	xx	Ossur Pro-Flex LP
102	42	Female	167	71	TT	Right	Infection	3	Test → Definitive	Pin lock	Gel liner	xx	Ossur Pro-Flex XC
103	38	Male	186	102	TT	Left	Infection	3	Definitive	Active Vacuum	Gel liner	xx	Ossur Pro-Flex XC
104	60	Male	178	84	TT	Left	Infection	3	Test → Definitive	Pin lock	Gel liner	xx	College Park Celsus
105	63	Male	176	120	TF	Bilateral	Trauma	3	Definitive	Passive suction	Gel liner	Ottobock C-leg	Ottobock Taleo
106	53	Female	162	68	TT	Right	Vascular	x	Definitive	Sleeve	Gel liner	xx	xx
107	44	Male	180	96	TT	Left	Trauma	3	Definitive	Active Vacuum	Gel liner	xx	Rush Rogue Evacuate
108	23	Male	172	70	TF	Left	Sarcoma	3	Test	Passive suction	Gel liner	Ossur Rheo XC	Ossur Pro-Flex XC
109	77	Male	188	63	KD	Right	Vascular	2	Definitive	Anatomical	Gel liner	Ossur OH5	Freedom Sierra
110	77	Female	152	79	TT	Left	Vascular	3	Test → Definitive	Pin lock	Gel liner	xx	Ottobock Restore
111	75	Male	184	95	TF	Right	Sarcoma	3	Definitive	Pin lock	Gel liner	Ottobock Genium	Ottobock Trias
112	60	Male	173	91	TT	Right	Sarcoma	3	Definitive	Active Vacuum	Gel liner	xx	Meridium; College park odyssey K3
114	35	Female	158	74	HD	Right	Sarcoma	3	Test	Anatomical	None	Ottobock 3R60-HD	Ottobock Terion K2
115	58	Male	167	79	TF	Right	Trauma	3	Definitive	Passive suction	Gel liner	Ottobock Genium X3	xx
116	63	Male	177	55	TT	Left	Infection	3	Definitive	Pin lock	Gel liner	xx	Ottobock Trias
117	69	Female	162	85	TT	Bilateral	Vascular	2	Test	Sleeve	Gel sock	xx	Ottobock Restore
120	22	Male	185	75	KD	Left	Trauma	3	Definitive	Anatomical	Gel liner	KX06	xx
121	75	Male	177	95	TF	Left	Infection	2	Definitive	Passive Suction	Gel liner	Ottobock 3R60	xx
122	52	Male	175	118	TT	Right	Infection	3	Definitive	Pin lock	Gel liner	xx	Rush Rampage LP
123	64	Male	190	90	TT	Bilateral	Vascular	3	Test → Definitive	Pin lock	Gel liner	xx	Ottobock Senator
124	69	Male	172	64	TT	Right	Vascular	3	Test	Pin lock	Gel liner	xx	Ossur Proflex LP

Level = Level of amputation. Side = Side of amputation. TT = Transtibial. TF = Transfemoral. KD = Knee Disarticulation. HD: Hip Disarticulation. X = participant's K-level was not reported. Socket = type of prosthetic socket. Knee = prosthetic knee. Foot = prosthetic foot. Test → Definitive = subject started data collection in a diagnostic socket and transitioned to a definitive socket during data collection. XX = prosthetic information is not relevant due to level of amputation or was not recorded.

Evaluation of Pipeline Step 1: Subject Identification and Tracking

The initial step of the gait analysis pipeline is to identify the person of interest in the frame to track for the duration of the video. If the tracking step is not working properly, the remaining steps of the pipeline will fail as well. A total of 270 level walking videos were included in the analysis. Of the 270 videos, none were discarded completely due to bounding box failure. However, 14 videos had to be processed with a different tracking method in order to correctly identify the person of interest throughout the duration of the video. In 263 of the 270 videos, the person of interest was identified in the frame for greater than 90 percent of the video.

Prosthetic Keypoint Quality compared to Keypoint Confidence

To determine if the computer-generated joint confidence values were accurate as a measure of 2D keypoint detection quality, the keypoint confidence values were compared to the prosthetic keypoint quality values. Figure 1A shows the relationship between annotated prosthetic keypoint quality and ankle keypoint confidence produced by the detection algorithm. A total of 186 videos were rated as a prosthetic keypoint quality 1, 144 videos were rated as a prosthetic keypoint quality 2, and 69 videos were rated as a prosthetic keypoint quality 3. The total number of ratings are greater than the 270 videos included as the bilateral subjects' videos were counted twice in order to include prosthetic keypoint quality values for each prosthetic limb. A one-way ANOVA was performed to compare the effect of keypoint quality on ankle keypoint confidence, and showed a significant difference between groups ($F(2, 405)=241.65$, $p=8e-70$). A post hoc Tukey test showed that all pairs of groups were significantly different ($p<0.001$). Figure 1B shows the relationship between prosthetic keypoint quality and knee keypoint confidence. The same number of videos were recorded for each prosthetic keypoint quality as above.

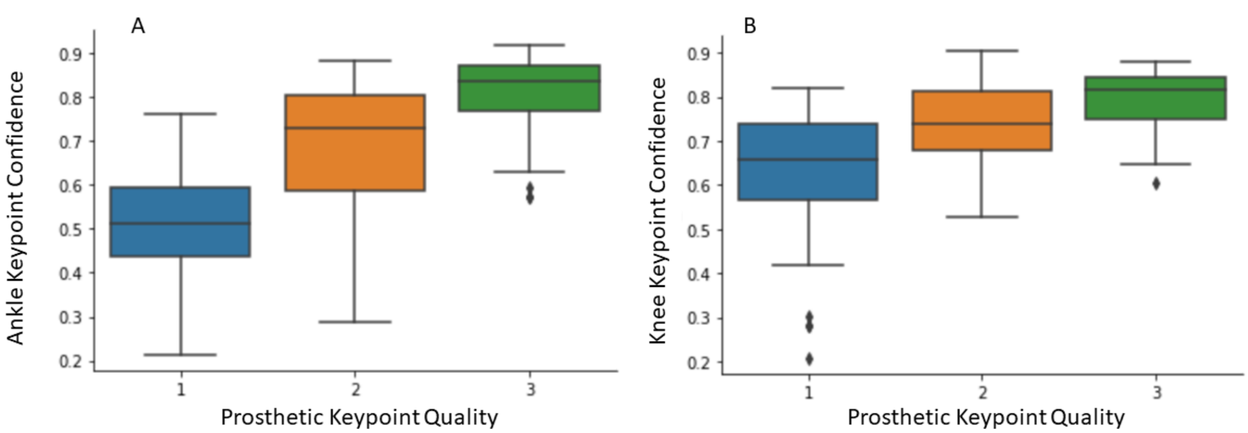

Figure 1: Prosthetic Keypoint Quality Compared to Keypoint Confidence. A: Prosthetic Keypoint Quality Compared to Ankle Keypoint Quality. B: Prosthetic Keypoint Quality Compared to Knee Keypoint Quality. Boxplot: Box identifies IQR (25-75 percentile), middle line identifies median, error bars indicate range, points outside the error bars indicate outliers.

Next, the trends for the ankle keypoint confidence and the knee keypoint confidence were compared to identify if one was a better summary statistic to evaluate the 2D joint tracking quality in a video. Average knee keypoint confidence was slightly higher than average ankle confidence for all annotated levels of prosthetic keypoint quality values, although this gap was most pronounced for prosthetic keypoint quality 1. Reviewing videos with the 2D keypoints superimposed also qualitatively confirmed that the ankle was typically the joint with the greatest tracking error. Because of this and that the ankle keypoint confidence shows a tighter correspondence to the annotated keypoint qualities (Fig 1A vs 1B), the ankle keypoint confidence was used as a summary statistic for tracking quality in following sections.

Evaluation of Pipeline Step 2: Factors that affect 2D keypoint identification

When manually annotating the videos, it was noted that the 2D keypoint quality appeared to be influenced by whether the prosthetic components were visible. In order to further confirm this observation, the ankle confidence values were compared when participants' prosthesis was visible and covered. Figure 2A shows the relationship between average ankle keypoint confidence with the prosthesis visible and covered. A total of 235 videos with the prosthesis visible and a total of 74 videos with the prosthesis covered were included. The total number of videos included in this figure is greater than the 270 total videos included in the study as videos of bilateral patients were included twice to account for both prosthetic limbs. The average ankle keypoint confidence associated with the prosthesis visible was 0.607, and the average ankle keypoint confidence associated with the prosthesis covered was 0.824. A two-way ANOVA was performed to analyze the effect of prosthetic visibility and subject ID on ankle keypoint confidence and showed a significant main effect of prosthetic visibility ($F(1, 375)=77.6$, $p=4.9e-17$). Figure 2B shows the ankle keypoint confidence for those participants whose data included videos of both conditions: prosthesis visible and covered. Every subject showed a higher average ankle keypoint confidence when their prosthesis was covered.

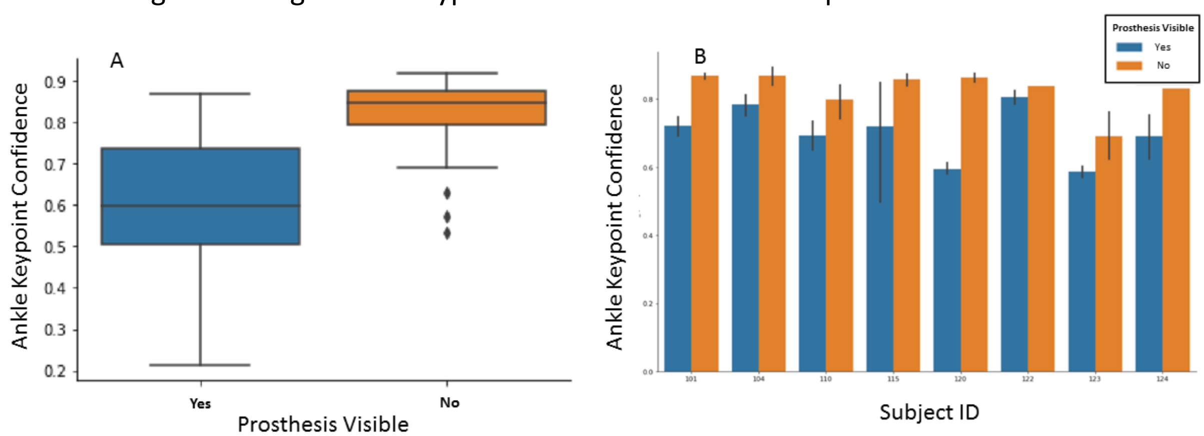

Figure 2: Relationship between Ankle Keypoint Confidence and Prosthesis Visible. A: Ankle Keypoint Confidence for both prosthesis visible conditions. Boxplot: Box identifies IQR (25-75 percentile), middle line identifies median, error bars indicate range, points outside the error bars indicate outliers. B: Ankle Keypoint Confidence for participants with data collected in both prosthesis visible conditions. Blue = prosthesis visible, Orange = prosthesis covered.

It was also subjectively noted that 2D keypoint tracking appeared to be worse for participants with more proximal or bilateral levels of amputation. To confirm this observation, the relationship between amputation level and ankle keypoint confidence was analyzed. Figure 3 shows the relationship between ankle keypoint confidence and level of amputation, further stratified for whether the prosthesis was visible. When the prosthesis was visible, average ankle keypoint confidence was greater for lower levels of amputation (transtibial > bilateral transtibial > transfemoral > bilateral transfemoral > hip disarticulation). A two-way ANOVA was performed to analyze the effect of prosthetic visibility and amputation level on ankle keypoint confidence which showed a significant interaction for both amputation level ($F(4,306)=667.3$, $p=1e-149$) and prosthetic visibility ($F(1,306)=126.5$, $p=8.5e-25$). A Tukey HSD posthoc test showed a significant pairwise difference between all prosthetic levels ($p<0.001$) other than between transfemoral and bilateral transtibial. When the prosthesis was covered, average ankle keypoint confidence was similar for transtibial and transfemoral participants. No data were collected of bilateral transfemoral or hip disarticulation patients with their prosthesis covered.

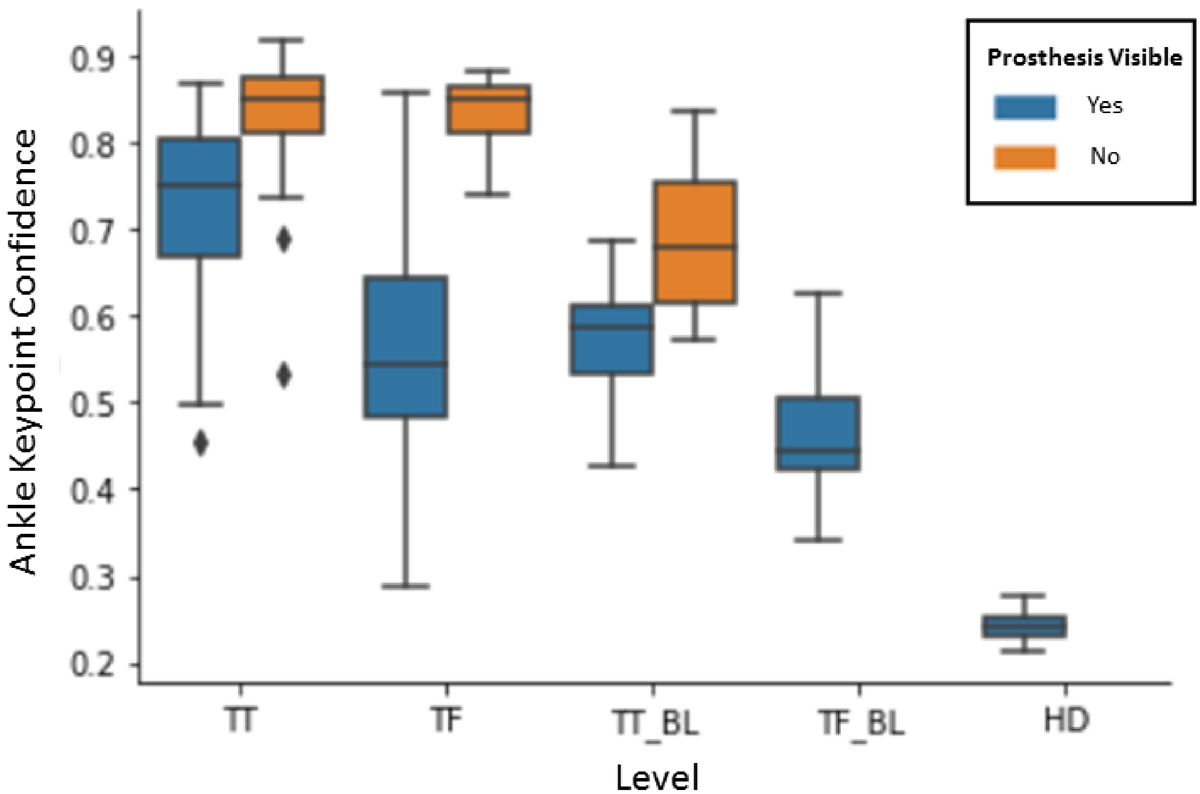

Figure 3: Relationship between Ankle Keypoint Confidence and Level of Amputation. Blue = prosthesis visible, Orange = prosthesis covered. TT = transtibial, TF = transfemoral, TT BL = bilateral transtibial, TF BL = bilateral transfemoral, HD = hip disarticulation. Boxplot: Box identifies IQR (25-75 percentile), middle line identifies median, error bars indicate range, points outside the error bars indicate outliers.

The type of prosthetic socket (definitive vs diagnostic) is another prosthetic related factor that was noted to potentially influence 2D keypoint detection. Therefore, this relationship was analyzed further. Figure 4 shows the relationship between ankle keypoint confidence and type of prosthetic socket for all individuals who had data recorded of them with both types of prosthetic sockets. All videos included in this figure were with pant legs up so that the prosthetic socket was visible. Average ankle keypoint confidence was higher for all subjects when they were utilizing their definitive prosthetic socket. A two-way ANOVA was performed to analyze the effect of subject ID and socket type on ankle keypoint confidence, and revealed a statistically significant difference between average ankle keypoint confidence and type of prosthetic socket ($F(1,79)=44.2$, $p=3.5e-9$), for subject ID ($F(3,79)=31.9$, $p=1.3e-13$), and the interaction of subject ID and socket type ($F(3,79)=3.6$, $p=1.7e-2$). Subject 110 had the greatest difference between the two conditions. This subject had a segment of pipe insulation foam covering the prosthetic pylon in their definitive socket condition. No other subjects had prosthetic covers.

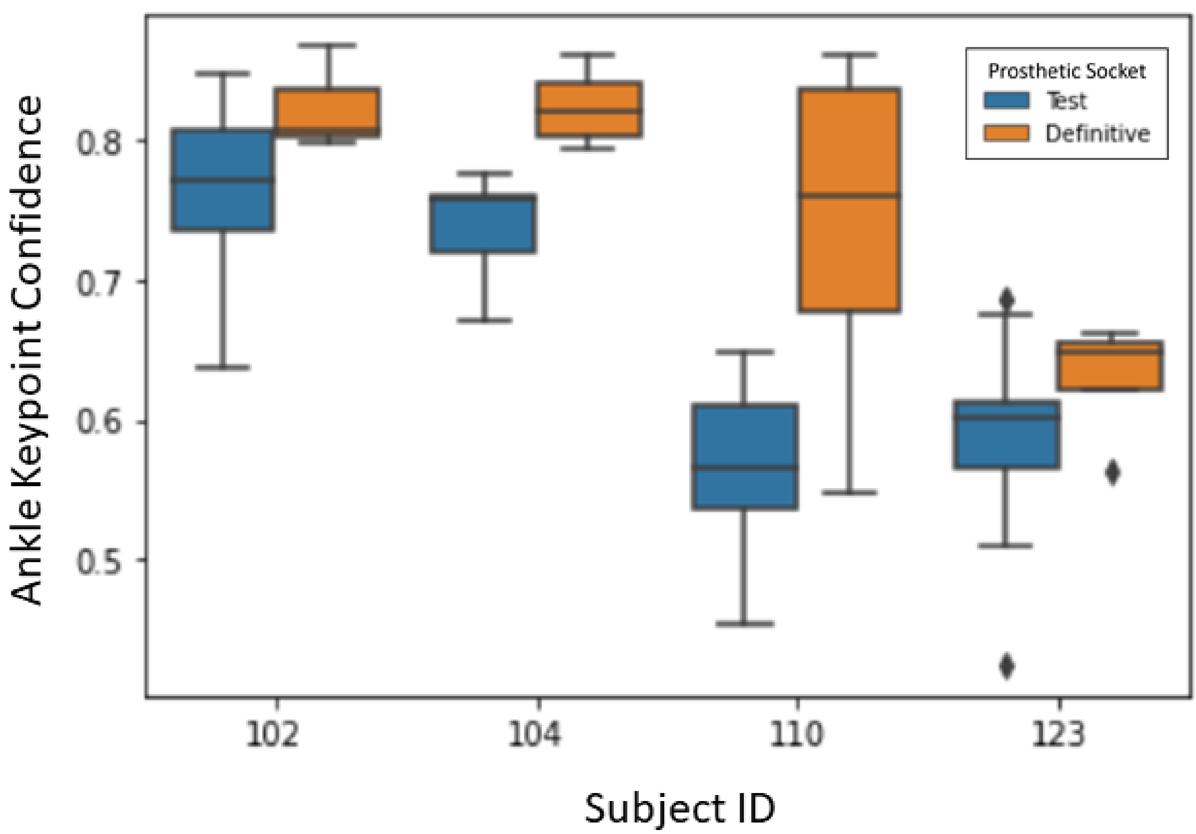

Figure 4: Relationship between Ankle Keypoint Confidence and Type of Prosthetic Socket. Blue = participant was utilizing a test/diagnostic prosthetic socket. Orange = participant was utilizing a definitive prosthetic socket. All videos were with the prosthesis visible. Boxplot: Box identifies IQR (25-75 percentile), middle line identifies median, error bars indicate range, points outside the error bars indicate outliers.

Videos were obtained in a variety of views from subjects including pure frontal plane, pure sagittal plane, and a mixture of both planes. The view was partially constrained by the room in the therapy areas and other patients or therapists present. In order to determine if the plane of view influenced 2D keypoint detection, the relationship between ankle keypoint confidence and plane of view was analyzed. Figure 5 shows the relationship between ankle keypoint confidence and plane of view the video was recorded in separated by whether the prosthesis was visible. When the prosthesis was visible, the average ankle keypoint confidence was slightly higher for frontal plane > sagittal plane > both planes. When the prosthesis was covered, the average ankle keypoint confidence between each group was very similar.

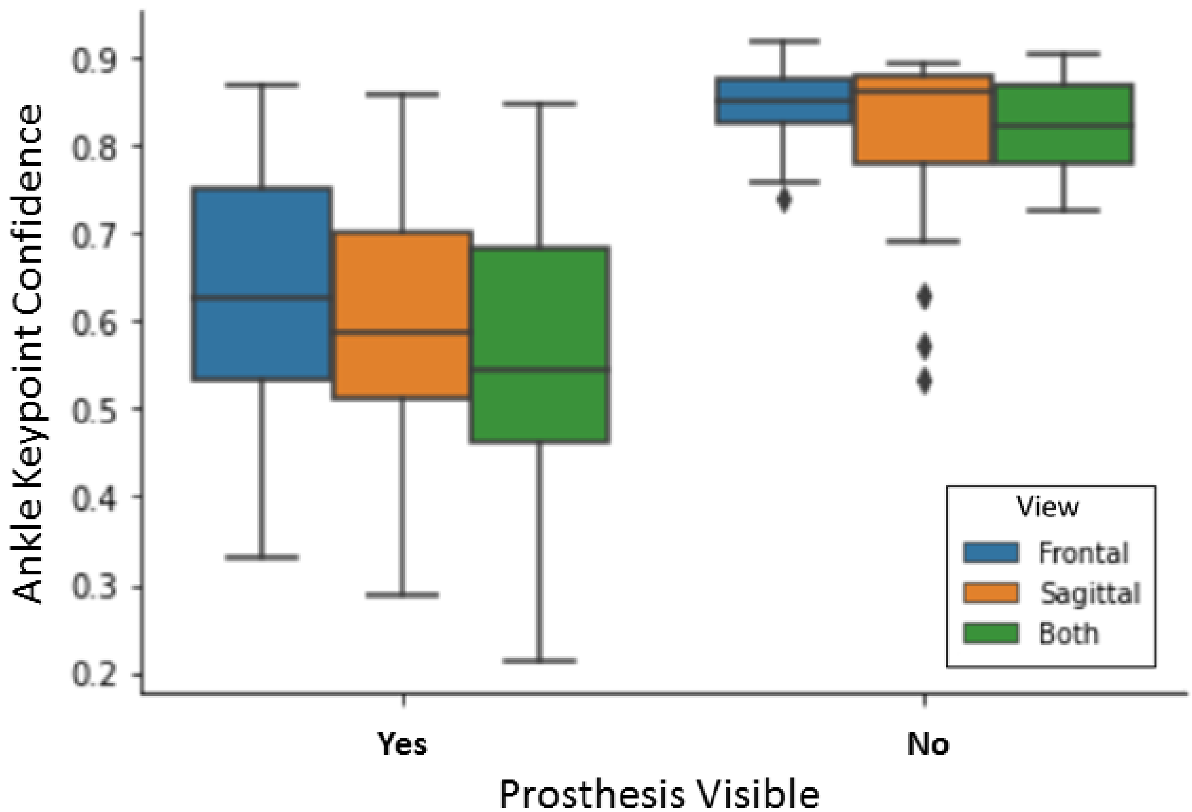

Figure 5: Relationship between Ankle Keypoint Confidence and Plane of View. Blue = frontal plane, Orange = sagittal plane, Green = both planes. Both indicates that the plane of view changed during the video or video was taken on an angle between both planes. Boxplot: Box identifies IQR (25-75 percentile), middle line identifies median, error bars indicate range, points outside the error bars indicate outliers.

Evaluation of Pipeline Step 3: Factors that affect gait transformer outputs

In order to accurately measure gait parameters over the duration of a video, the gait transformer must first accurately detect when the person of interest is ambulating. As the data included in the results of this study were limited to level walking activities, the participants were ambulating for most of the video. However, the length of each video and therefore duration of ambulation varied between participants and videos. In addition, the total number of videos for each subject varied due to their clinical schedule and activities they were working on during clinical visits. Table 2 shows the amount of walking detected for each participant. There were several individuals (103, 105, 114, 117, 121 and 123) in which walking was detected less than 10% of the time. Walking was complete not detected in two individuals (109 and 124). Of those eight individuals: one had bilateral transfemoral amputations, two had bilateral transtibial amputations, one had a hip disarticulation amputation, two had a transfemoral or knee disarticulation amputation and two had a transtibial amputation.

Table 2: Detecting Walking by Participant.

Subject ID	Level	Total Videos	Walking Segments	Walking Frames	Total Frames	Fraction Frames
107	TT	5	11	3839	4517	84.99
115	TF	10	31	9033	12858	70.25
122	TT	5	21	6827	9857	69.26
112	TT	5	16	6315	9615	65.68
104	TT	16	64	27533	42410	64.92
108	TF	13	132	26874	55456	48.46
116	TT	21	181	50897	106803	47.66
101	TT	8	13	3679	7823	47.03
106	TT	15	59	15642	53203	29.40
120	KD	5	11	2503	9329	26.83
102	TT	20	83	21999	85167	25.83
110	TT	43	94	26217	127886	20.50
111	TF	28	64	12011	60568	19.83
103	TT	2	1	188	2097	8.97
117	B TT	10	9	1719	21697	7.92
123	B TT	19	20	3722	90734	4.10
121	TF	17	13	1796	72380	2.48
114	HD	6	4	583	34840	1.67
105	B TF	23	8	990	63321	1.56

Level = level of amputation. Total videos = total number of level walking videos. Walking segments = number of times the system detecting walking. Walking frames = number of frames the system detecting walking. Total Frames = total frames collected on participant. Fraction Frames = % of frames that walking was detected.

To determine if 2D keypoint detection was affecting the outputs of the gait transformer, the relationship between ankle keypoint confidence and walking detection was analyzed. Figure 6A shows the relationship between ankle keypoint confidence and fraction of frames that walking was detected. While there is no direct linear relationship between ankle keypoint confidence and fraction of frames, it does appear that when ankle keypoint confidence is high, the fraction of frames walking is detecting tends to be higher. And when the ankle keypoint confidence is low, fraction of frames that walking is detected is also low.

Kalman Error is another measure used to determine the quality of the gait transformer outputs. Therefore, the relationship between ankle keypoint confidence and Kalman error was further explored to determine if 2D keypoint detection was affecting the quality of the gait transformer outputs. Figure 6B shows the relationship between ankle keypoint confidence and Kalman error. Again, there is no direct linear relationship between ankle keypoint confidence and Kalman error. However, it appears that once the ankle tracking quality drops below a certain level the Kalman error increases and consequently the fraction of frames in which walking is detected tends to decrease. Thus, 2D keypoint tracking appears to have an impact on the quality of the gait transformer outputs.

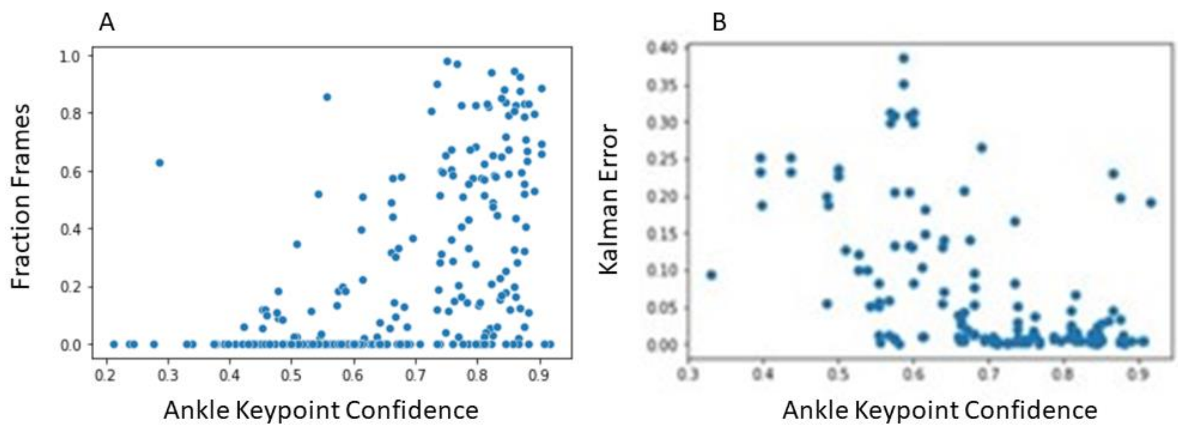

Figure 6: Relationship between Ankle Keypoint Confidence and Gait Transformer Outputs. A: Ankle Keypoint Confidence and Fraction of Frames. B: Ankle Keypoint Confidence and Kalman Error.

As it was previously determined that ankle keypoint confidence and walking detection appear to be affected by amputation level, it was thought that the Kalman error might also be affected by level of amputation. If so, it could indicate that the accuracy of the gait parameter values output by the gait transformer may be affected by amputation level. Therefore, this relationship was further investigated. Figure 7 shows the relationship between Kalman Error and Level of Amputation separated by whether the prosthesis was visible. Average Kalman error was different between levels of amputation. With higher levels of amputation having greater Kalman error (Bilateral TT > Bilateral TF > TF > TT). In addition, for the transtibial and transfemoral groups, the Kalman error was higher when the prosthesis was visible.

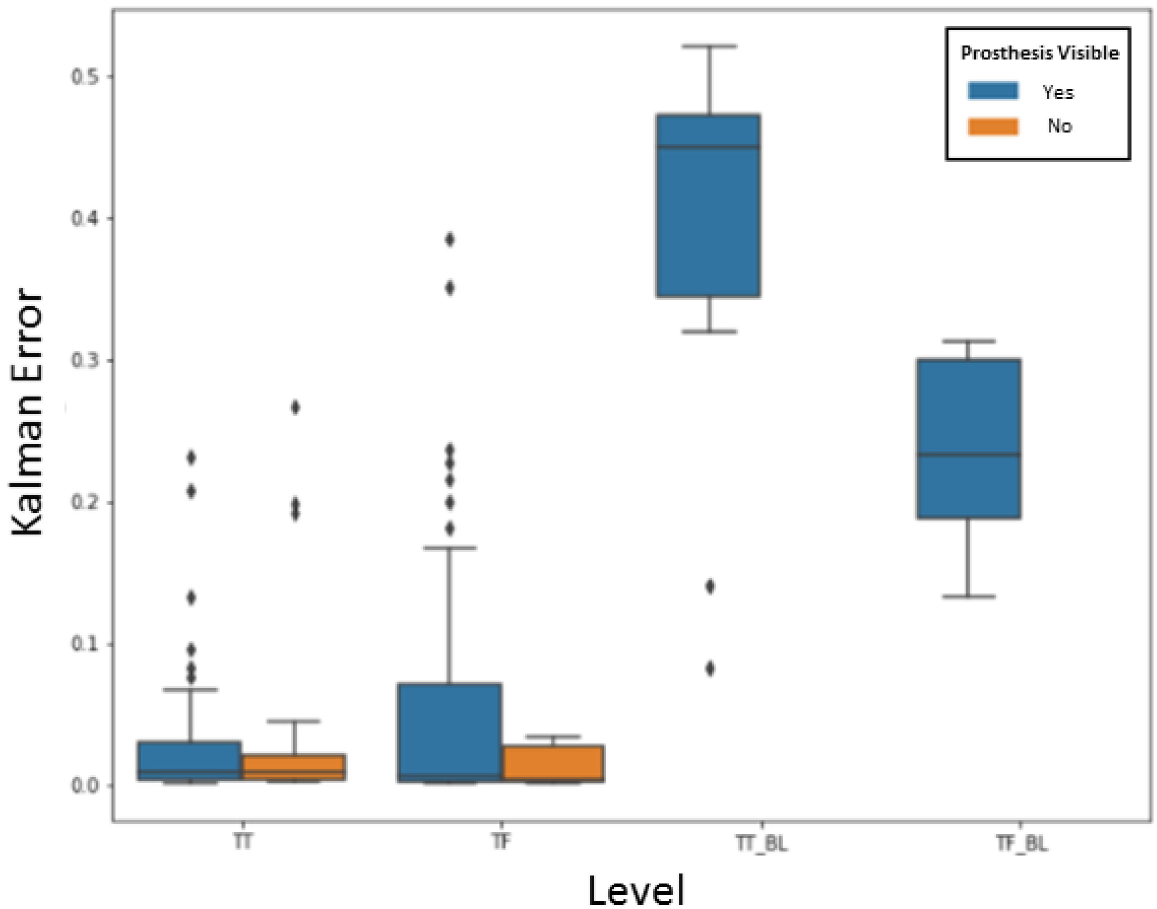

Figure 7: Relationship between Kalman Error and Level of Amputation. Blue = prosthesis visible, Orange = prosthesis covered, TT = transtibial, TF = transfemoral, TT_BL = bilateral transtibial, TF_BL = bilateral transfemoral. Boxplot: Box identifies IQR (25-75 percentile), middle line identifies median, error bars indicate range, points outside the error bars indicate outliers.

Evaluation of Pipeline Step 3: Accuracy of Gait Transformer Measures

The gait transformer has been trained to output several gait parameters and averages these parameters over the time it detects the individual walking. Once the gait transformer is accurately detecting walking, the next step is to access the accuracy of the gait parameters that it measures. While the gait transformer outputs several parameters, in this study, the accuracy of the gait transformer velocity and cadence were analyzed.

Figure 8A shows the relationship between the gait transformer velocity and the annotated velocity. Gait transformer velocity was calculated as the average velocity when the system was detecting walking and the individual was ambulating between the two black lines. Annotated velocity was calculated by manually timing 10-meter walk with black tape marks on the ground in the clinical area. A total of 86 videos were included in this figure, with 186 walking segments from 15 different subjects. Only videos where the individual was ambulating without stopping for a break between the two black lines, and the system was detecting walking at some point during that time were included in this figure. The gait transformer velocity measure was relatively accurate compared to the annotated velocity (R^2 = 0.740, r = 0.894). In general, when individuals were ambulating at slower speeds (<1.0 m/s), the gait transformer tended to overestimate their velocity and when individuals were ambulating at faster speeds (>1.0 m/s), the gait transformer tended to underestimate their velocity.

Figure 8B shows the relationship between the gait transformer cadence and the cadence measured by the wearable sensors. The gait transformer cadence was calculated as the average cadence when the system was detecting walking in steps/minute. The sensor cadence was measured as steps/minute calculated from the wearable sensor attached to the prosthesis when the gait transformer was detecting walking. Only walking segments that were at least 300 frames long were included. A total of 61 videos were included in this figure, with 135 walking segments from 15 different subjects. The gait transformer cadence was relatively accurate compared to the sensor cadence (R^2 = 0.694, r = 0.843). The correlation was very tight other than a single outlier where the gait transformer estimated the cadence as zero that may be negatively affecting our results.

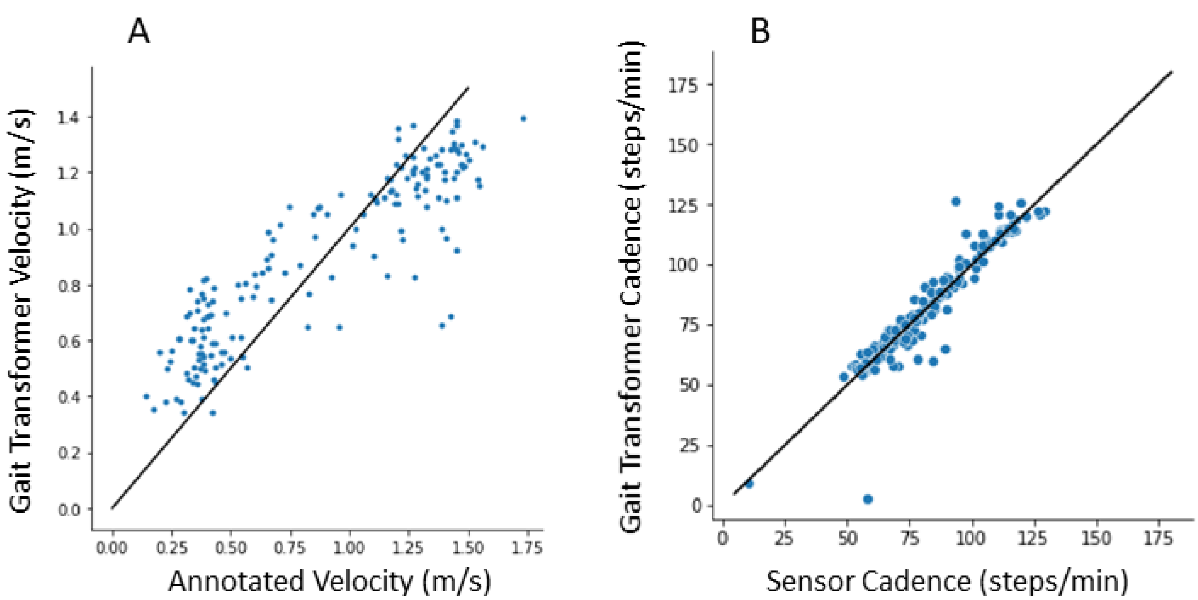

Figure 8: Accuracy of Gait Transformer Measures. A: Gait Transformer Velocity compared to Annotated Velocity. B: Gait Transformer Cadence compared to Sensor Cadence.

Longitudinal assessment in physical therapy: demonstrate potential use as outcome measure

One of the main goals for the application of this technology is to produce a system to improve quantitative, longitudinal characterization of gait in a clinical setting. Figure 9 demonstrates the potential to use the gait analysis system as an outcome measure to detect longitudinal change in prosthetic users over the course of several physical therapy sessions. Four participants who had data collected at a minimum of three physical therapy sessions over the course of at least five weeks were selected to assess for change. Sensor cadence was averaged for the entire duration of level walking videos recorded that day. Gait transformer cadence was averaged over the times the system was detecting walking. Annotated velocity was calculated using the times the individual crossed the black lines on the ground in the clinical area. Gait transformer velocity was averaged over the times the system was detecting walking.

In general, there was a good correspondence between the gait transformer measures and the values measured from the wearable sensors and annotation of walking speed. The gait transformer cadence values were well correlated with the sensor cadence values for each participant and the gait transformer cadence was sensitive to change that the sensor cadence measured between days. However, there were several notable exceptions when the gait transformer outputs were inaccurate. For subject 102 (Figure 10A), no gait transformer outputs were measured the first week, indicating that the system did not detect walking at all during those videos. In addition, several subjects saw a difference between the gait transformer velocity and the annotated velocity. For subjects 102, 104 and 116 the gait transformer tended to overestimate the velocity compared to the annotated velocity. While for subject 108, the gait transformer tended to underestimate the velocity. For subjects 104 and 108, while the gait transformer velocity values were slightly different from the annotated values, both values appeared to still show the same general trend in change in velocity. However, for subjects 102 and 116, the gait transformer reported changes in velocity while the annotated velocity remained more stable.

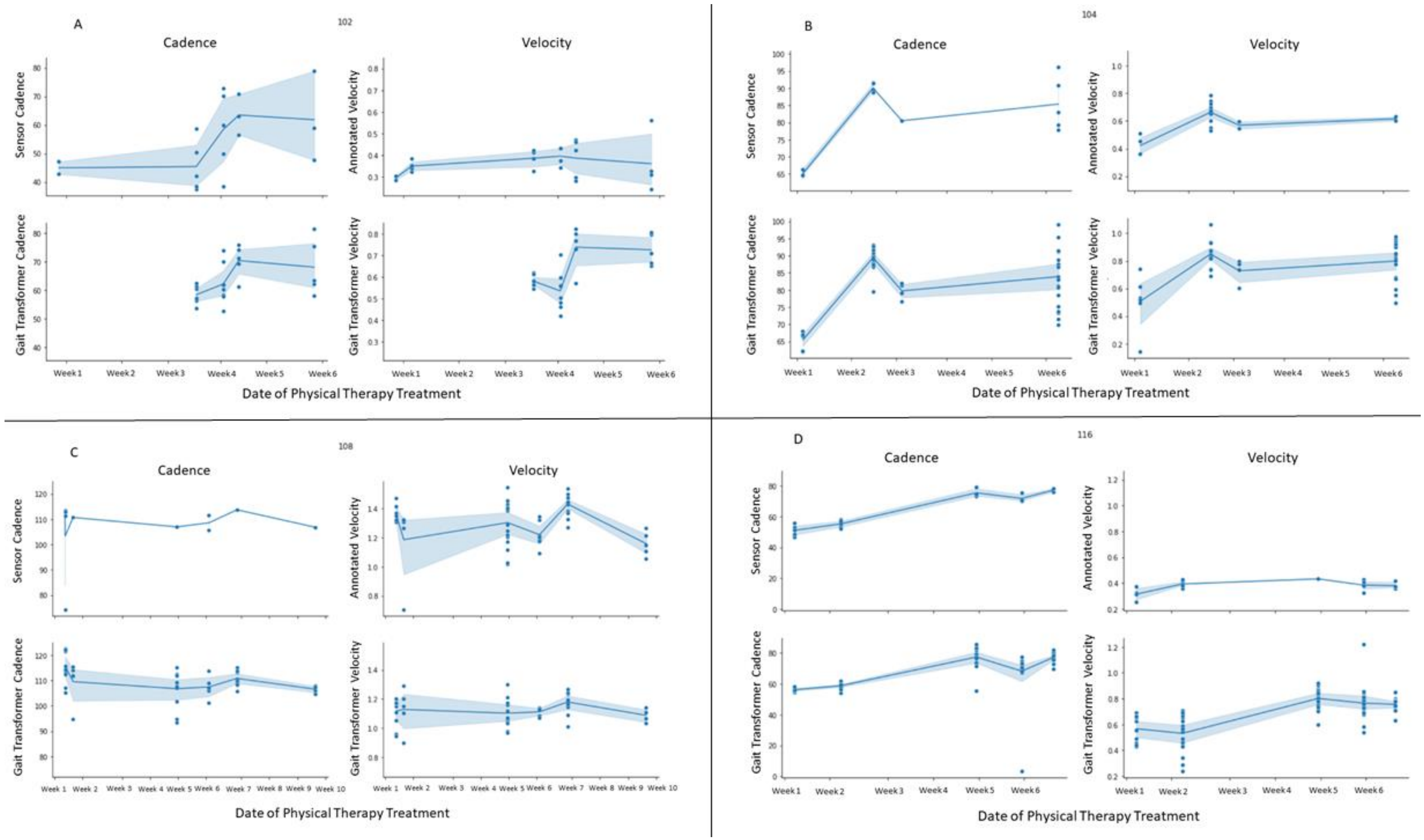

Figure 9: Longitudinal Assessment in Physical Therapy: Demonstration of potential use as an outcome measure. Each figure shows the comparison between cadence and velocity over the course of several therapy sessions with date on the x-axis and cadence/velocity on the y-axis. Date ranges on the x-axis indicated the week of physical therapy treatment the data was collected. Left: sensor cadence compared to gait transformer cadence. Right: annotated velocity compared to gait transformer velocity. A: Subject 102. B: Subject 104. C: Subject 108. D: Subject 116. Dots represent measurements recorded from that date, line is the average, shaded lines represent the error.

Discussion

In order to evaluate using video-based gait analysis with our custom gait analysis pipeline to track longitudinal outcomes for prosthetic users, we performed an in-depth analysis of the performance of several components including Subject Identification and Tracking, 2D Keypoint Identification, and Gait Transformer Outputs. This revealed numerous reasons for optimism but also highlighted several limitations.

Evaluation of Pipeline Step 1: Subject Identification and Tracking

Identifying the person of interest and tracking them throughout the duration of the video is a step in the pipeline that worked well and does not appear to be a limiting factor. However, it is still not perfect and 4% of videos did require processing with an alternative algorithm with different idiosyncrasies. Problems that necessitated this were most commonly identity switches between the prosthetic user and therapist, likely exacerbated by the need for them to remain in close proximity for safety reasons. This also sometimes occurred when both the prosthetic user and therapist left the frame, which is sometimes inevitable when acquiring video in a therapy gym. All videos were able to be processed and identify the person of interest. In nearly all the videos, the person of interest was identified in the frame for greater than 90 percent of the time. However, 100 percent identification may not be possible in all videos as there were times in some of the videos when the subject left the frame completely as mentioned above.

Evaluation of Pipeline Step 2: Factors that affect 2D keypoint identification

The algorithms' ability to correctly identify the 2D joint location is fundamental to creating an accurate 3D model of the individual ambulating. Since the pre-trained algorithms had only been trained on able-bodied individuals, one would expect that visual limb differences between prosthetic users and able-bodied individuals might result in poor 2D keypoint generalization. When the prosthesis was showing, average ankle keypoint confidence was lower (Figure 2). In addition, for all participants who had data collected in both prosthesis visible conditions, ankle keypoint confidence was lower for all individuals when their prosthesis was visible indicating that the pre-trained algorithms are struggling to identify the 2D keypoints when the prosthetic limb is showing.

As the level of amputation becomes more proximal, the limb differences between the prosthetic device and an able-bodied limb increase. Therefore, one would assume that the higher the level of amputation the worse the 2D keypoint tracking would be. Results confirmed that when the prosthesis was visible, the higher the level of amputation, the lower the ankle keypoint confidence (Figure 3). Furthermore, when the entire prosthesis was covered there was no difference in ankle keypoint confidence between transtibial and transfemoral users. Therefore, it is assumed that as the level of amputation increases, the more distal ankle joint is harder for the computer to estimate, as there is more prosthetic componentry showing that is visually different in shape from typical able-bodied limbs, and therefore it is less confident in the location of the ankle joint.

The type of prosthetic socket (definitive vs diagnostic) is a prosthetic related factor that could affect 2D keypoint detection when the prosthesis is visible as the materials used in the two types of sockets may be identified differently by the algorithms. Though not the case with the study population, individuals with definitive sockets sometimes have prosthetic covers that cover the pylon and make the prosthesis appear more like an anatomical leg, which would likely improve 2D keypoint detection. Of the four subjects with data in both socket conditions, only one had some sort of cover. This subject had a piece of pipe foam covering the prosthetic pylon that increased the size of the shank without providing the shape of an anatomic calf. When analyzing the ankle keypoint confidence of these individuals with their prosthesis visible, there was a difference in ankle keypoint confidence between having a definitive or diagnostic socket. The subject with the foam cover had the greatest difference in average ankle keypoint confidence between the diagnostic and definitive condition. This indicates that the 2D keypoint detection does appear to be affected by the type of prosthetic socket and presence of a foam cover. However, our sample size for this analysis was limited to four subjects and therefore may not be large enough to make assumptions to a larger number of individuals. Typically, diagnostic sockets are made of clear plastic and may have various different shapes near the distal end of the socket as the connection between the prosthetic socket and distal prosthetic componentry varies between individual based on the shape of their residual limb, their prosthetic alignment and the suspension mechanism that is built into the socket. In addition, diagnostic sockets are typically reinforced with fiberglass at the distal end, creating a color transition in the socket. While definitive sockets are typically a solid color and have a more streamlined shape at the distal end. These different shape outlines and color variations may have an effect on the algorithms ability to detect 2D keypoints. Furthermore, it appears that any factor that makes the limb appear more like an anatomical limb (lower levels of amputation, prosthesis covered, and definitive prosthetic socket) increases the quality of 2D keypoint detection.

For the algorithms to be able to correctly identify the joint locations, one would assume that the joints would likely have to be visible in the frame. Therefore, it was thought that the plane of view the video was taken in might affect 2D keypoint detection. However, our results showed that there was no difference in 2D keypoint detection between the planes of view (Figure 5). This is an important finding, as the goal is to use the gait analysis system in a clinical setting. Often space is limited in clinical settings and therefore the plane of view that the video may be recorded in could be constrained by the space available. In these situations, it is likely more important for the person recording the video to position themselves in a location where they can keep the person of interest in the frame the entire time rather than worrying about what plane the video is being recorded in.

Other factors that were analyzed and did not appear to have an effect on 2D keypoint detection include presence of an assistive device and walking velocity. Presence of an assistive device is a factor that could potentially influence 2D keypoint detection, as there is more in the image for the algorithm to detect and determine the location of the joints. In addition, often the assistive device may be obstructing the view of the joints and therefore affecting 2D keypoint detection. Clinically, individuals who ambulate with an assistive device tend to do so because they have a lower level of mobility or need the device for balance security. Therefore, they likely ambulate differently than most able-bodied individuals. As the sample included in this study was a convenience sample and data were collected during normal clinical visits with the intention of not interfering with clinical practice, presence of an assistive device was not controlled for. Therefore, we did not have enough data of individual subjects ambulating with and without an assistive device to determine if the assistive device was affecting 2D keypoint detection or if other factors related to the individual were affecting 2D keypoint detection. Walking velocity did not appear to have a direct influence on 2D keypoint detection. This is likely because the 2D keypoints are detected in each frame of the video and regardless of how fast the individual is ambulating, the algorithms should be able to detect the joints in each frame if the images are clear. However, walking velocity may have an effect on the ability of the gait transformer to detect walking.

Evaluation of Pipeline Step 3: Factors that affect gait transformer outputs

There are two variables that can be used to evaluate the gait transformer outputs in order to determine if is functioning properly for prosthetic users. The first is that the system is detecting walking (Fraction Frames). The second is Kalman Error, an ad hoc measure to determine the error associated with the gait transformer output. The system's ability to detect walking is fundamental for accurate measurements because the values of interest are subsequently averaged over the time the system detects walking. Therefore, if it is not detecting walking accurately, relevant data will not be included in the final measurements. As the videos were typically only recorded when the participants were ambulating, we can assume that the participant was ambulating the entire duration of the video. Walking was detected more than 50% of the time for just five of the twenty-one participants. The system was limited in detecting walking for several individuals and two participants were not detected walking at all. This may indicate that the walking detection criteria used was too conservative, or that there are other prosthetic related factors affecting the systems' ability to detect walking such as 2D keypoint detection. The individuals with higher levels of amputation and the bilateral participants tended to be the ones that the system did not detect walking well.

As these individuals also tended to have lower ankle keypoint confidence, the relationship between ankle keypoint confidence and the gait transformer outputs was analyzed (Figure 6). While there is not a direct relationship between ankle keypoint confidence and Fraction of Frames or Kalman Error, 2D keypoint quality seems to be related to the quality of the gait transformer outputs. It appears that once the ankle tracking quality drops below a certain level the Kalman error increases and consequently the fraction of frames walking is detected tends to decrease. This indicates that 2D keypoint tracking likely has an impact on the quality of the gait transformer outputs. Furthermore, individuals with higher levels of amputation, tended to have high Kalman error (Figure 7). Indicating that the gait transformer is not performing as well for individuals with higher levels of amputation.

Our results have shown that for individuals with higher levels of amputation, 2D keypoint quality was significantly lower than lower levels of amputation, which may have affected the system's ability to detect walking in these individuals. Clinically, bilateral prosthetic users and those with higher levels of amputation tend to ambulate with more asymmetries and gait deviations compared to able bodied individuals [3]. In addition to the prosthetic factors explored, this may have affected the gait transformer's ability to understand how they are ambulating as it has been trained on individuals with intact limbs. If the initial step of 2D keypoint detection is inaccurate, the gait transformer outputs will likely be inaccurate as well. However, as the gait transformer was trained on non-amputees, the asymmetries present in prosthetic gait, specifically among those with higher levels of amputation or bilateral amputations, may have affected the gait transformers outputs. More work is needed to improve 2D keypoint detection and walking detection in prosthetic users. Once these steps are improved, additional analysis and work may be needed to improve the gait transformer's ability to understand some of the asymmetries present in prosthetic gait. We have identified a potential solution to improve 2D keypoint detection. While the process is quite time-consuming, initial results show some promising improvement in 2D keypoint detection of the prosthetic limb. For more details please see Image 3 in the Appendix.

Evaluation of Pipeline Step 3: Accuracy of Gait Transformer Measures and Demonstration of Potential Use as an Outcome Measure

When the system is detecting walking accurately, the accuracy of the various output parameters also needs to be evaluated. Our results show that when the system is detecting walking properly, it is fairly accurate at measuring velocity and cadence in prosthetic users (Figure 8). However, the system tended to overestimate velocity for individuals who were ambulating slowly, and underestimate velocity for individuals who were ambulating quickly. This inaccurate measure of velocity may affect the ability to use the system as an outcome measure to detect changes in velocity over time as seen in Figure 9. When using the gait analysis system to detect change in prosthetic users over the course of several physical therapy sessions, a difference was seen between the annotated velocity and gait transformer velocity for a few individuals. It is possible that the system failed to detect walking accurately and the gait transformer velocity used for longitudinal assessment was only a small portion of the time the individual was ambulating. However, it is also possible that the amount of change in the individual's velocity was small enough that it is within the gait transformer's velocity error range and therefore the gait transformer does not accurately detect this change in velocity.

The system's ability to measure cadence accurately did not appear to be affected by the speed the individual was ambulating at, as seen in figure 8B. In addition, when using the system to detect changes in cadence over the course of several physical therapy sessions, there was no difference between the gait transformer cadence and the sensor cadence as seen in Figure 9. Therefore, in its current state, the gait transformer appears to be able to measure changes in cadence accurately. However, additional work is needed to improve the gait transformers velocity measurement. As velocity is a measure that is easily measured with current outcome measures, next steps will look more into other quantitative gait data to measure such as step length, step width, kinematics and kinetics.

Clinical Implications of Results of this study

The goal of this study was to determine the possibility and identify challenges of using currently available markerless pose estimation techniques in conjunction with a custom gait transformer to perform gait analysis in prosthetic users in a clinical setting. The results demonstrate that when using pre-trained pose estimation algorithms for gait analysis on prosthetic users, the best results will likely be accomplished on individuals who appear more like those with intact limbs. Therefore, best results will be accomplished on individuals with lower levels of amputation and when the prosthesis is covered by clothing, as 2D keypoint detection is better for those individuals. If the prosthesis is visible, better results will be achieved with individuals utilizing a definitive socket. However, clinically the standard of practice is to observe a prosthetic user ambulating with their prosthesis showing so the prosthetist can assess what is occurring at the prosthesis and make any necessary adjustments. Therefore, in order to make this technology more applicable for prosthetists in a clinical setting, additional work is needed to improve 2D keypoint detection when the prosthetic limb is visible. Furthermore, additional work is needed to determine if specific socket colors, materials or cosmetic covers affect results as well.

In its current state, the custom gait analysis pipeline used in this study appears to be fairly accurate for measuring walking velocity and cadence. Furthermore, the gait transformer appears to work better on individuals with lower levels of amputation. A possible reason for this is that individuals with higher levels of amputation tend to ambulate with more asymmetries and gait deviations compared individuals with intact limbs. Therefore, in order to use the gait transformer on individuals with higher levels of amputation, the transformer may need to be trained to better understand prosthetic gait and different prosthetic componentry. Future work will also look to measure additional gait parameters.

In addition, as these algorithmic based deep learning techniques have not been widely used in rehabilitation, it will be important for future work to carefully annotate and evaluate outputs and results to ensure accuracy at each stage of data processing.

Limitations

The main limitation to this study was the small sample size. Convenience sampling was used to enroll subjects and the number of individuals in each group may not be equal. Therefore, it is difficult to generalize the results to larger populations. In addition, several factors analyzed were not necessarily controlled for as this study was conducted in a clinical setting with the goal being to collect data without interfering with routine clinical practice. These factors include whether the prosthesis was visible, presence of an assistive device, plane of view, and type of prosthetic socket. When possible, results were analyzed for individuals who had data collected with both conditions for each of these factors. However, there may be other factors that were not controlled for when analyzing certain relationships that could affect the results.

Implications of the use of a gait analysis system in a clinical setting

While there is still work needed to improve the gait analysis pipeline to increase the accuracy of its outputs, the results of this study demonstrate that it is possible to measure quantitative gait data from a cell phone video in a clinical setting. The ability to measure quantitative gait data from easily obtainable video could have many potential uses in rehabilitation and research. These uses may include more detailed outcome measurement, mobile gait analysis for use in research studies to assess community ambulation, dynamic alignment tool for prosthetists to assist with observational gait analysis and alignment, as well as for ankle foot orthosis tuning. While the results from video-based gait characterization may not be as accurate as traditional gold standard measures, if they allow for measurement in a clinical or community setting where useable gait data was not previously easy to obtain, results may be very useful despite these inaccuracies. As future work continues to improve the results of these techniques, the question of how accurate is sufficient for use in a clinical/community setting will need to be addressed.

Future directions

There are many potential applications and future directions in this area of research. Next steps specific to use in prosthetic users will first be to improve 2D keypoint detection and walking detection for higher levels of amputation. Once those steps are improved, additional training to improve the gait transformer's ability to understand some of the gait asymmetries apparent in prosthetic gait may be warranted. Additional parameters to be measured from video analysis include additional spatial-temporal, kinematic and kinetic parameters. Furthermore, the outputs of the gait transformer will likely need to be validated against current gold standard measurement techniques to determine the accuracy of its outputs. Future work will likely include testing and training of the algorithms on additional patient populations for use in a clinical setting in order to increase the amount of people the gait analysis system can be used with. Once the system is properly trained, and outputting valid and accurate gait data, work will need to be done to determine how this system can be used specifically to aide in clinical decision-making.

Conclusions

The results of this study demonstrated that it is possible to measure quantitative gait data from video obtained with a cell phone camera in a clinical setting. However, there are many factors that affect the ability to use pre-trained pose estimation algorithms on prosthetic users. Best results were obtained for individuals who appear more like those with intact limbs (i.e. lower levels of amputation and when the prosthesis was covered by clothing). Differences in limb and gait characteristics between prosthetic users and the individuals that the pre-trained algorithms and the gait transformer were trained on appear to be affecting results. Additional work is needed to improve the system's ability to work properly with prosthetic users. However, with improvements in the ability to locate 2D joints and accurately detect walking, this system shows the potential to be used quickly and easily in a clinical setting with minimal equipment or training required.

Appendix

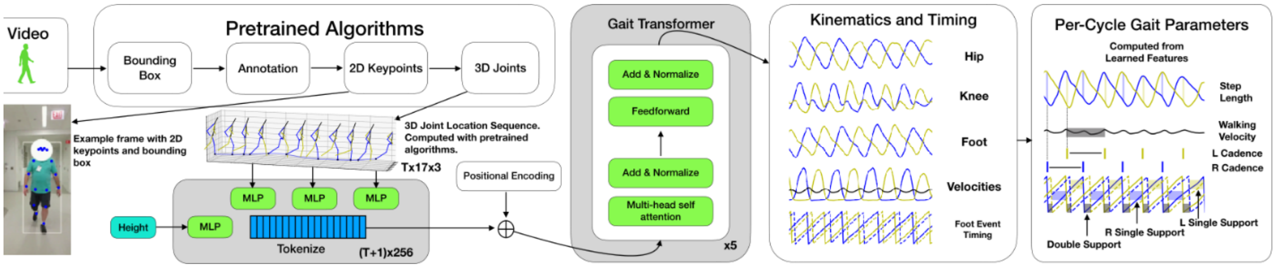

Figure 10: Overview of the custom gait analysis pipeline.

Potential Solution to Improve 2D Keypoint Detection

We have identified a potential solution to improve 2D keypoint detection for the prosthetic limb. The process involves manual annotation of each video with a software called Deep Lab Cut. To annotate a video, 20 frames are chosen and the prosthetic joints are manually annotated in each frame. The software then takes those manually annotated frames and applies the joint locations to the entire video. While the process is quite time-consuming, initial results show promising improvement in 2D keypoint detection as seen in Image 3.

Image 3: Example of improvement in 2D Keypoint Detection with Deep Lab Cut Annotation. A: 2D Keypoint Detection with Pre-Trained Algorithm. B: 2D Keypoint Detection with Deep Lab Cut Annotation. C: 2D Keypoint Detection with Pre-Trained Algorithm. D: 2D Keypoint Detection with Deep Lab Cut Annotation. E: 2D Keypoint Detection with Pre-Trained Algorithm. F: 2D Keypoint Detection with Deep Lab Cut Annotation.

References

1. Ziegler-Graham K, MacKenzie EJ, Ephraim PL, Travison TG, Brookmeyer R. Estimating the Prevalence of Limb Loss in the United States: 2005 to 2050. *Archives of Physical Medicine and Rehabilitation*. 2008;89(3):422-429. doi:10.1016/J.APMR.2007.11.005
2. Vital and Health Statistics From the CENTERS FOR DISEASE CONTROL AND PREVENTION / National Center for Health Statistics. Published online 1998.
3. Queen RM, Orendurff M. Amputee Gait : Normal and Abnormal. In: *Atlas of Amputations and Limb Deficiencies Surgical, Prosthetic, and Rehabilitation Principles*. Vol 1. 4th ed. ; 2016:69-77.
4. Brinkmann, John T; Stevens PM. Clinical Considerations of Observational Gait Analysis. In: *Atlas of Amputations and Limb Deficiencies Surgical, Prosthetic, and Rehabilitation Principles*. 4th ed. ; 2016:81-95.
5. Mueller MJ, Salsich GB, Bastian AJ. Differences in the gait characteristics of people with diabetes and transmetatarsal amputation compared with age-matched controls. *Gait and Posture*. 1998;7(3):200-206. doi:10.1016/S0966-6362(98)00015-0
6. Klöpfer-Krämer I, Brand A, Wackerle H, Müßig J, Kröger I, Augat P. Gait analysis-Available platforms for outcome assessment R. *Injury*. 2020;51(2):90-96. doi:10.1016/j.injury.2019.11.011
7. SR K, YY D, AH H. Cross-validation of a portable, six-degree-of-freedom load cell for use in lower-limb prosthetics research. *Journal of biomechanics*. 2014;47(6):1542-1547. doi:10.1016/J.JBIOMECH.2014.01.048
8. Jonkergouw N, Prins MR, Van Der Wurff P, et al. Dynamic alignment using external socket reaction moments in trans-tibial amputees. Published online 2018. doi:10.1016/j.gaitpost.2018.11.004
9. Jonkergouw N, Prins MR, Buis AWP, Van Der Wurff P. The effect of alignment changes on unilateral transtibial amputee's gait: A systematic review. *PLoS ONE*. 2016;11(12):1-19. doi:10.1371/journal.pone.0167466
10. Rusaw D, Ramstrand N. Motion-analysis studies of transtibial prosthesis users: A systematic review. *Prosthetics and Orthotics International*. 2011;35(1):8-19. doi:10.1177/0309364610393060
11. Sagawa Y, Turcot K, Armand S, Thevenon A, Vuillerme N, Watelain E. Biomechanics and physiological parameters during gait in lower-limb amputees: A systematic review. *Gait & Posture*. 2011;33(4):511-526. doi:10.1016/J.GAITPOST.2011.02.003
12. Kent J, Franklyn-Miller A. Biomechanical models in the study of lower limb amputee kinematics: A review. *Prosthetics and Orthotics International*. 2011;35(2):124-139. doi:10.1177/0309364611407677
13. Kaluf B, Stevens P. Outcome Measures in Lower Limb Prosthetics. In: *Atlas of Amputations and Limb Deficiencies Surgical, Prosthetic, and Rehabilitation Principles*. 4th ed. ; 2016:645-662.
14. Hawkins EJ, Riddick W. Reliability, Validity, and Responsiveness of Clinical Performance-Based Outcome Measures of Walking for Individuals With Lower Limb Amputations: A Systematic Review. *Physical therapy*. 2018;98(12):1037-1045. doi:10.1093/ptj/pzy107

15. Hellsten T, Karlsson J, Shamsuzzaman M, Pulkkis G. The Potential of Computer Vision-Based Marker-Less Human Motion Analysis for Rehabilitation. *Rehabilitation Process and Outcome*. 2021;10:117957272110223. doi:10.1177/11795727211022330
16. Dang Q, Yin J, Wang B, Zheng W. *Deep Learning Based 2D Human Pose Estimation: A Survey Deep Learning Based 2D Human Pose Estimation: A Survey Deep Learning Based 2D Human Pose Estimation: A Survey*. Vol 24.
17. Kidziński Ł, Yang B, Hicks JL, Rajagopal A, Delp SL, Schwartz MH. Deep neural networks enable quantitative movement analysis using single-camera videos. *Nature Communications*. 2020;11(1):1-10. doi:10.1038/s41467-020-17807-z
18. Lu M, Poston K, Pfefferbaum A, et al. Vision-based Estimation of MDS-UPDRS Gait Scores for Assessing Parkinson's Disease Motor Severity. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2020;12263 LNCS:637-647.
19. James Cotton R. Kinematic Tracking of Rehabilitation Patients with Markerless Pose Estimation Fused with Wearable Inertial Sensors. In: *Proceedings - 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2020*. Institute of Electrical and Electronics Engineers Inc.; 2020:508-514. doi:10.1109/FG47880.2020.00092
20. Cotton RJ, Rogers J. Wearable monitoring of joint angle and muscle activity. *IEEE International Conference on Rehabilitation Robotics*. 2019;2019-June:258-263. doi:10.1109/ICORR.2019.8779538
21. Cotton RJ, Rogers J. Wearable monitoring of joint angle and muscle activity. *IEEE International Conference on Rehabilitation Robotics*. 2019;2019-June:258-263. doi:10.1109/ICORR.2019.8779538
22. Cotton RJ, Mcclerklin E, Cimorelli A, Patel A, Karakostas T. Transforming Gait : Video-Based Spatiotemporal Gait Analysis. *In Submission*. Published online 2022.
23. Wojke N, Bewley A, Paulus D. Simple Online and Realtime Tracking with a Deep Association Metric. *Proceedings - International Conference on Image Processing, ICIP*. 2017;2017-September:3645-3649. doi:10.48550/arxiv.1703.07402
24. Wu J, Cao J, Song L, Wang Y, Yang M, Yuan J. Track to Detect and Segment: An Online Multi-Object Tracker. Published online March 16, 2021. doi:10.48550/arxiv.2103.08808
25. Contributors Mmp. GitHub - open-mmlab/mmpose: OpenMMLab Pose Estimation Toolbox and Benchmark. Accessed March 14, 2022. https://github.com/open-mmlab/mmpose
26. Sun K, Xiao B, Liu D, Wang J. Deep High-Resolution Representation Learning for Human Pose Estimation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2019;2019-June:5686-5696. doi:10.48550/arxiv.1902.09212
27. Zhang F, Zhu X, Dai H, Ye M, Zhu C. Distribution-Aware Coordinate Representation for Human Pose Estimation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Published online October 14, 2019:7091-7100. doi:10.48550/arxiv.1910.06278
28. Lin TY, Maire M, Belongie S, et al. Microsoft COCO: Common Objects in Context. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and*

Lecture Notes in Bioinformatics). 2014;8693 LNCS(PART 5):740-755. doi:10.1007/978-3-319-10602-1_48

29. Liu J, Rojas J, Li Y, et al. A Graph Attention Spatio-temporal Convolutional Network for 3D Human Pose Estimation in Video. Published online March 11, 2020:3374-3380. doi:10.48550/arxiv.2003.14179
30. Rauch HE, Tung F, Striebel CT. Maximum likelihood estimates of linear dynamic systems. *https://doi.org/102514/33166*. 1965;3(8):1445-1450. doi:10.2514/3.3166