Comparison of Automated Event Detection Algorithms in Pathological Gait

Dustin A. Bruening
Brigham Young University, dabruening@byu.edu

Sarah Trager Ridge

Follow this and additional works at: https://scholarsarchive.byu.edu/facpub

Part of the Exercise Science Commons

Original Publication Citation

BYU ScholarsArchive Citation
https://scholarsarchive.byu.edu/facpub/1627

This Peer-Reviewed Article is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Faculty Publications by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
Automated event detection algorithms in pathological gait

Dustin A. Bruening, Sarah Trager Ridge

Abstract

Accurate automated event detection is important in increasing the efficiency and utility of instrumented gait analysis. Published automated event detection algorithms, however, have had limited testing on pathological populations, particularly those where force measurements are not available or reliable. In this study we first postulated robust definitions of gait events that were subsequently used to compare kinematic based event detection algorithms across difficult pathologies. We hypothesized that algorithm accuracy would vary by gait pattern, and that accurate event detection could be accomplished by first visually classifying the gait pattern, and subsequently choosing the most appropriate algorithm. Nine published kinematic event detection algorithms were applied to an existing instrumented pediatric gait database (primarily cerebral palsy pathologies), that were categorized into 4 visually distinct gait patterns. More than 750 total events were manually rated and these events were used as a gold standard for comparison to each algorithm. Results suggested that for foot strike events, algorithm choice was dependent on whether the foot's motion in terminal swing was more horizontal or vertical. For horizontal foot motion in swing, algorithms that used horizontal position, resultant sagittal plane velocity, or horizontal acceleration signals were most robust; while for vertical foot motion, resultant sagittal velocity or vertical acceleration excelled. For toe off events, horizontal position or resultant sagittal plane velocity performed the best across all groups. We also tuned the resultant sagittal plane velocity signal to walking speed to create an algorithm that can be used for all groups and in real time.

Keywords: Instrumented gait analysis; Gait cycle; Gait classification or pattern; Event detection; Automation
Introduction

Pathologies that impair the neuromuscular system are frequently treated with the assistance of instrumented gait analysis [1]; however, this technology can be costly both in equipment and manpower. One of the most time-consuming processes in instrumented gait analysis is event detection. Most commonly, the gait cycle is broken into stance and swing phases, separated by the events of foot strike (FS) and toe off (TO) [2] and [3]. Individual walking patterns can be quite variable, and in our experience with pediatric pathological gait, automated event detection algorithms often fail, necessitating time consuming manual inspection. In addition to increasing post-processing efficiency, accurate automation of event detection may also be essential for some real-time biofeedback applications, such as gait re-training [4].

The gold standard for automated event detection is generally assumed to be the onset and termination of foot-to-ground contact, determined by a force or pressure measuring device. These devices, however, are often limited or unavailable. There are also inherent limitations in assuming a force measurement as a gold standard for pathological populations, as some subjects for example slide or drag their feet in swing phase, creating false force thresholds. We postulate that a more comprehensive definition of gait events, covering pathologies as well as healthy subjects, might be described using the following Boolean logic:

\[ \text{FS} = \text{foot contacts the ground AND foot forward progression is arrested} \]

\[ \text{TO} = \text{toe leaves the ground OR toe starts forward progression} \]

These definitions include both kinetic and kinematic components, but it may be possible to substitute kinematic criteria for the former. Kinematic-only methods are generally more versatile, and as such have received substantial attention in the literature, with more than a dozen published algorithms, most showing good accuracy in normal gait [5], [6], [7], [8], [9], [10], [11], [12] and [13]. Testing on pathological gait, however, has been limited to small samples of mild pathologies [9], [11] and [12], has lacked comparisons with other algorithms [10] and [13], and all have relied on force data as a gold standard [6], [9], [10], [11], [12] and [13]. The purpose of the present study was to compare the accuracy of published kinematic-only event detection algorithms among several types of more challenging clinical gait patterns. It was hypothesized that algorithm accuracy would vary by gait pattern, so that accurate event detection could be accomplished by first visually classifying the gait pattern, and subsequently choosing the most appropriate algorithm.

Methods

This was a retrospective study utilizing an existing instrumented pediatric gait database (comprised primarily of cerebral palsy pathologies). Experienced technicians first identified
numerous subjects for which event detection was particularly challenging. These subjects were subsequently grouped by experienced researchers and clinicians into four distinct lower extremity movement classifications using physician and therapist diagnostic notations along with visual inspection of videos and model based graphics:

**Equinus:** FS is made by the mid or forefoot instead of the heel, common among equinus or jump knee classifications [14]. The foot sometimes also has increased plantarflexion at TO.

**Slide/drag:** This is two related classifications, one focused on FS (slide) and one on TO (drag). For slide, the foot has little clearance in swing (e.g. stiff knee or foot drop) and FS is approached horizontally, often with ground contact prior to FS. For drag, the toes or medial forefoot drag across or near the ground as swing phase is initiated. While a slide FS and a drag TO sometimes occur in the same subject, for this study we used separate subjects for each group.

**Steppage:** Subjects had a marching type gait, characterized by more vertical motion at (primarily) FS. This is usually confined to subjects that use a walker or other assistive device. TO is more variable in this group, with some exhibiting near normal patterns, while others had a slight vertical pattern or even a drag pattern.

**Miscellaneous:** A mix of challenging gait patterns that did not easily fit the previous categories. These included very slow walking speeds and transverse plane (rotational) abnormalities.

Ten subjects from each category were analyzed, for a total of 50 subjects (slide and drag categories employed separate subjects). All steppage group subjects used assistive devices as did one equinus, three drag, and four miscellaneous subjects. Only one limb from each subject was used (generally the more involved of the two).

A lower extremity model had been previously implemented for each subject in Visual 3D software (C-Motion Inc., Germantown MD, USA) for clinical use, and lower extremity joint kinematics had been calculated. Of relevance to this study, all data was collected at 120 Hz using approximately 40, 6-mm diameter retro-reflective markers. Marker trajectories were filtered at 6 Hz using a low pass Butterworth filter (4th order, dual pass). Visual 3D was also used for visualization of marker trajectories and foot segments for event rating.

Between 7 and 16 events were rated for each subject, across 3–6 walking trials, for a total of 404 FS and 373 TO events. All events had been previously rated for clinical use; however, for the purposes of this study, an experienced rater reviewed all previous events and manually checked or re-rated each event, using the above definitions (Eqs. (1) and (2)) and including force plate data as a guide when available (33 of the 50 subjects had force plate data for at least 4 events). FS was considered the first frame of stance and TO the last frame of stance. These manually rated events were considered the gold standard for the subsequent comparisons. To help measure manual rating agreement, two additional experienced raters independently determined the same events for a random subset consisting of 2 subjects from each group with a total of 159 events.
Inter-rater reliability was measured as the difference in frames between these additional raters (raters 2 and 3) and the gold standard (rater 1).

We identified nine kinematic-based algorithms from peer reviewed literature that appeared promising for event detection in these challenging pathologies. The chosen algorithms are listed by the primary author's last name for organizational convenience, along with a brief description of the algorithm criteria (note that unless otherwise specified, the description refers to a marker or markers on the foot):

- **Zeni** [13]: Peak anterior (FS) and posterior (TO) position relative to the pelvis.
- **Desailly** [6]: Peak positive (FS) and negative (TO) high pass filter of horizontal position.
- **Ghoussayni** [7]: Sagittal plane descending (FS) and ascending (TO) velocity threshold.
- **O'Connor** [11]: Peak negative (FS) and positive (TO) vertical velocity.
- **Salazar-Torres** [12]: Horizontal/sagittal velocity descending (FS) and ascending (TO) threshold.
- **Hreljac** [8]: Peak vertical (FS) deceleration and horizontal (TO) acceleration.
- **Hsue** [9]: Peak horizontal (FS) deceleration.
- **DeAsha** [5]: Peak contralateral hip extension (FS).
- **Jasciewicz** [10]: Peak foot angular velocity (plantarflexion) (TO) and zero crossing (FS).

Heel, forefoot (2nd metatarsal head), and sacrum marker positions, along with foot angular velocity and contralateral hip joint angles were exported from Visual 3D. Custom LabView (National Instruments Co, Austin, TX, USA) and Matlab (Mathworks Inc., Natick, MA, USA) software were then used to program all algorithms. Most algorithms consist of two separable parts: (1) – the choice of signal to use (e.g. heel acceleration, hip extension, etc.), and (2) – the method of finding the relevant peak or threshold in that signal that corresponds to the desired event. In many cases the signal contained multiple peaks, and determining the correct peak proved to be quite difficult, particularly using original algorithm methodology. We made the choice in this study to manually identify difficult peaks; in other words, we assumed that the correct peak or threshold could be identified, and comparisons were made only on part 1. Additionally, two notable deviations were made to Desailly and Ghoussayni. The cutoff frequency recommended for TO in Desailly (1.1 × cadence) was extremely sensitive to subject cadence and signal endpoint conditioning, and we lowered it to match that recommended for FS (0.5 × cadence). The velocity threshold recommended by Ghoussayni was also found to be too low, and we increased it an order of magnitude from 5 cm/s to 50 cm/s for all subjects. Also note that in the Equinus group, the forefoot marker was used in place of the heel marker for FS. The forefoot marker was used to determine TO in all groups.
After event calculation, the gold standard manually determined event frames were then subtracted from the calculated event frames, so that negative differences represented early events, while positive differences represented late events.

**Results**

Subset reliability testing (Fig. 1) showed inter-rater differences that were generally less than the algorithm errors (compare with Fig. 2 and Fig. 3). No meaningful biases were noted for the gold standard rater with the exception of a possible slight early bias for FS of the *Slide* group (Fig. 1). If all three raters are compared in pair wise fashion, 84% of all pair wise ratings are within 2 frames and 92% within 3 frames.

![Fig. 1. Inter-rater reliability.](image)

*Fig. 1. Inter-rater reliability. Difference between the gold standard ratings (rater 1) and those of the subset raters 2 and 3 (rater 2–rater 1 and rater 3–rater 1). Reliability ratings were done on a random subset of two subjects from each group (total of 159 events). Each point is a single event, with raters separated by colors (grayscale when colors are not available).*
Fig. 2. Foot strike (FS). Difference between algorithm determined FS events and gold standard manually determined FS events, organized by group (Equinus, Slide, Steppage, and Miscellaneous). Each point represents a single event, while colors are used to group events belonging to the same subject (grayscale when colors are not available). Positive differences are events occurring later than the gold standard.
Fig. 3. Toe off (TO). Difference between algorithm determined TO events and gold standard manually determined TO events, organized by group (Equinus, Slide, Steppage, and Miscellaneous). Each point represents a single event, while colors are used to group events belonging to the same subject (grayscale when colors are not available). Positive differences are events occurring later than the gold standard.

Based on the reliability testing as well as a review of the algorithm results (Fig. 2 and Fig. 3), a 4 frame window (33 ms) was chosen for convenience in presenting and discussing the algorithm accuracy and precision; however, all results are displayed in the figures.

FS events (Fig. 2): in the Equinus group, four algorithms (Zeni, Desailly, Hreljac, Ghoussayni) identified at least 98% of FS events within four frames of the manual events. Zeni and Desailly were generally in agreement, consistently identifying FS a few frames early, while Hreljac and Ghoussayni were a few frames late. Similar results were found in the Slide group (swapping
Hsue for Hreljac), with Zeni, Desailly, Hsue, and Ghoussayni all identifying at least 96% of events within four frames of the manual events. Hsue was consistently a few frames early. Salazar-Torres also performed well, with 94% of events within four frames. For the Steppage group, Hreljac identified 95% of all FS events within four frames. Ghoussayni also performed well for most subjects (86% within four frames). Hreljac again had the highest percentage of events within four frames (79%) in the Miscellaneous group, notably with no outliers (i.e. high precision). Ghoussayni and Desailly, the next highest, had just 65% of events within four frames.

TO events (Fig. 3): For Equinus TO, Ghoussayni, Zeni, and Desailly identified TO within four frames 91%, 85%, and 75% of the time, respectively. Zeni and Desailly were consistently a few frames late. These same three algorithms also had the highest percentages within four frames in the Drag group at 89%, 72%, and 70%, respectively, as well as in the Steppage group, at 63%, 69%, and 68%, respectively. Desailly and Zeni were the most accurate in the Miscellaneous group with 84% of events within four frames for both.

**Discussion**

In this study we first visually classified challenging pathological gait into event patterns, and then compared the ability of kinematic based algorithms to accurately identify FS and TO. Algorithm performance did vary by group as hypothesized, yet there were also trends across groups with some algorithms consistently outperforming others, suggesting stronger inherent associations between those signals and the event definitions (Eqs. (1) and (2)).

DeAsha and Jascewicz both focused on model based joint or segment measures (hip angle and foot angular velocity, respectively). While in healthy gait local changes in these signals occur close to FS and TO, these signals are indirect measures of the events, explaining their generally poor performance in the pathologies of this study. For example, hip flexion was often more of an extended plateau than a single peak, and the location of the local maximum varied greatly across the plateau.

The algorithms of Zeni and Desailly gave results generally within one or two frames of each other. Both algorithms employed horizontal foot position signals, and it is therefore not surprising that they performed well for the more horizontal gait of the Slide group. Conversely, they performed poorly for FS of the vertical Steppage group, as many of these subjects reached local maximums in horizontal position during midswing, before lowering the foot to the ground. Both algorithms were among the best at identifying TO across all groups, and foot movement at TO appears to be consistently much more horizontal than at FS. We should also note that while Desailly recommends using a high pass filter, the effect of the filter at low frequencies is to simply transform the non stationary, step-like horizontal foot position signal into a stationary one (oscillating about zero). The same effect can alternately be accomplished by simply fitting a line to the signal and subtracting this line from each value. Zeni may have a slight advantage in robustness due to its simpler implementation, but Desailly does not require a marker on the pelvis. Both Desailly and Zeni tested their algorithms on some subjects with pathologies,
reporting good accuracy. The included pathologies were likely much milder than those in this study, and match the good accuracy we also showed for certain groups.

Three algorithms utilized first derivatives of foot position. Ghoussayni included both horizontal and vertical velocity components, and the resultant magnitude appeared to be fairly robust for FS and TO across all groups (see also additional analysis and commentary below). Salazar-Torres took a somewhat similar approach but chose a ratio between the signals. This worked well for some events (e.g. Slide FS); however, this success is partially due to the added manual threshold inspections that were performed in this study and greatly understates the possibility of errors in this extremely variable signal. O’Connor chose only the vertical component, for which we found generally poor performance and potential difficulty in peak identification. A position threshold was suggested to avoid false peak detections, but we found that this threshold would require subject specific tuning to be an effective aid in pathological populations.

The timing of the vertical velocity peak is also at odds with its derivative, the vertical acceleration peak used by Hreljac. For most groups, peak vertical velocity typically occurred several frames or more before the peak acceleration and the manually identified FS. Hreljac identified FS effectively for several groups, with the exception of the Slide group, where horizontal movement dominated. Here, the horizontal component of acceleration suggested by Hsue was more accurate. Acceleration did not appear to be a particularly effective signal for many of the TO events across all groups, as pathological gait often has a gradual TO transition that is not easily captured in an acceleration peak.

Across all groups and events, the sagittal plane resultant velocity signal (Ghoussayni) appeared to us to show the most potential versatility, for several reasons: First, the resultant velocity captures both horizontal and vertical movement in a Boolean-like relationship similar to Eqs. (1) and (2): as the velocity decreases in terminal swing, FS is identified only when BOTH components drop below a certain value; conversely, as the velocity increases, TO is identified when EITHER component rises above the threshold. Second, a velocity threshold can be tuned for subject specificity; and third, it may be effective in real-time applications, without waiting for the post-event frames needed for peak identification. We subsequently tuned the threshold with the current data to see if we could improve algorithm accuracy. We chose to tune it to walking speed, which can often be determined a priori and showed a reasonably strong correlation with the resultant velocity at the manual events ($R = 0.78$ for FS, $R = 0.77$ for TO). The threshold could then be calculated as a simple function of walking speed:

$$FS \text{ threshold } = 0.78 \times Walk \text{ Speed}$$

$$TO \text{ threshold } = 0.66 \times Walk \text{ Speed}$$

Tuning the algorithm with the above equations, we were able to modestly increase the accuracy of Ghoussayni to capture FS within 4 frames in all but 4 subjects (three of whom used assistive devices). TO was also slightly improved, but inaccuracies were still present in many subjects. It appeared that most of these inaccuracies could be greatly improved by using a marker on the
Hallux, as premature movement by the metatarsal heads was common among nearly all of the false thresholds. Unfortunately, this could not be tested with the current data.

Marker location may affect algorithm performance. The use of a hallux marker would likely increase TO accuracy across groups and algorithms, but presents some problems in placement. For Drag subjects, a marker on the medial aspect of the hallux may be useful [15], while for those with rotational deviations a marker on the interphalangeal joint [16] or hallux nail [17] may help. For FS, the use of a forefoot marker in the Equinus group appeared to make very little difference for most algorithms, with the exception of Hreljac, which was consistently improved by a few frames as the peak forefoot deceleration occurred prior to peak heel deceleration.

We relied on manual event identification as the best available gold standard for the clinical populations in this study. In many cases, gait transitions occur gradually and there is a range of frames within which identification of a single event frame is subjective. Our subset reliability testing suggests that this range was generally no more than a few frames, and less than the algorithm errors. Compared with the other two raters, the gold standard rater showed a possible early bias for FS of the Slide group (Fig. 1). The practical effect of this bias would simply be a slight increase in the accuracy of Ghoussayni and a slight decrease in the accuracy of Zeni, Desailly, and Hsue. Slide FS is gradual, and limb loading appears to begin before the foot's forward movement is completely arrested. If it is deemed important that some of this loading be presented in the stance phase instead of swing, as our gold standard rater appears to have done, this suggests a possible minor limitation in Eq. (1).

Summary and application

Our results can be applied to routine clinical practice, by identifying and implementing the most appropriate algorithm(s) for each specific gait pattern. While our subjects had primarily cerebral palsy, and represented extremely challenging cases, the results are instructive from a general standpoint in identifying those signals that are most robust or most closely related to the inherent event definitions. For FS, our recommendation includes sagittal resultant velocity (Ghoussayni), horizontal position (Zeni/Desailly), or vertical/horizontal acceleration (Hreljac/Hsue), depending on whether terminal swing is more horizontal or vertical. For TO, horizontal position (Zeni/Desailly) and sagittal velocity (Ghoussayni) appeared most accurate across all groups. TO accuracy will also likely be increased using a hallux marker. It may also be possible to calculate events using multiple algorithms; e.g. if the differences among these are substantial the computer can flag an operator to perform a manual check. If a single algorithm is desired for all subjects, or real-time event detection is needed, we think that Ghoussayni shows the most promise, particularly when tuned to walking speed or perhaps even to a kinetic measure (e.g. [18]). We hope that this study will also lead to the development of new algorithms with increased robustness in identifying events accurately and consistently across all pathologies.
References


