AWARD NUMBER: W81XWH-15-1-0445

TITLE: Automated Assessment of Postural Stability (AAPS)

PRINCIPAL INVESTIGATOR: Iyad Obeid, PhD

CONTRACTING ORGANIZATION: Temple University
Philadelphia PA 19122

REPORT DATE: October 2017

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE October 2017	**2. REPORT TYPE** Annual	**3. DATES COVERED** 15Sep2016 - 14Sep2017

4. TITLE AND SUBTITLE
Automated Assessment of Postural Stability (AAPS)

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-15-1-0445

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Iyad Obeid, PhD; Carole Tucker, PhD; Alessandro Napoli, PhD; Stephen Glass, PhD.

E-Mail: iobeid@temple.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

TEMPLE UNIVERSITY-OF THE COMMONWEALTH SY
1801 N BROAD ST
PHILADELPHIA PA 19122-6003

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Year 2 of this project was devoted to expanding the AAPS system capabilities and functionalities. The system has been optimized in terms of reliability and performance. Over the year, the AAPS has been extensively used for balance screening and data collection in different environments, such as: fitness centers, office spaces, outdoor settings and laboratories. The large number of subjects and data collection sessions performed over the year have led to a substantial software improvement and optimization, namely, multiple software bugs have been identified and solved, and the GUI has been redesigned and fine-tuned to ensure optimal performance and usability. We have also made substantial progress towards developing an extended version of the system that is capable of quantifying dynamic movement as opposed to static balance poses.

15. SUBJECT TERMS
motion tracking, balance assessment, Microsoft Kinect, concussion assessment

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** USAMRMC
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**c. THIS PAGE** Unclassified	Unclassified	24	**19b. TELEPHONE NUMBER** *(include area code)*

Table of Contents

Introduction

Year 2 of this project was devoted to expanding the AAPS system capabilities and functionalities. The system has been optimized in terms of reliability and performance. Over the year, the AAPS has been extensively used for balance screening and data collection in different environments, such as: fitness centers, office spaces, outdoor settings and laboratories. The large number of subjects and data collection sessions performed over the year have led to a substantial software improvement and optimization, namely, multiple software bugs have been identified and solved, and the GUI has been redesigned and fine-tuned to ensure optimal performance and usability. We have also made substantial progress towards developing an extended version of the system that is capable of quantifying dynamic movement as opposed to static balance poses.

Keywords

motion tracking, balance assessment, Microsoft Kinect, concussion assessment

Accomplishments

What were the major goals and objectives of the project?

The purpose of this project is to create a portable system for assessing balance in armed forces personnel that can be administered in the field with minimal training. Although there are many reasons for assessing an individual's sense of balance, our project focuses on balance deficits caused by concussion, traumatic brain injury, and musculoskeletal injury, since these are especially relevant to fitness for duty. Our deliverable will be a stand-alone system comprising a Microsoft Kinect motion tracking system and a dedicated laptop personal computer running custom software for data acquisition and analysis. The system is called the Automated Assessment of Postural Stability, or AAPS.

The project is designed around four Specific Aims, or goals:

1. Develop Baseline AAPS System
2. AAPS Calibration and Baseline Evaluation
3. AAPS Field Evaluation
4. Develop Expanded xAAPS Test

What was accomplished under these goals?

Accomplishment 1: AAPS Error Detection

The AAPS main innovation is its balance error detection algorithm that has been designed to acquire data from a Microsoft Kinect® sensor and convert them into clinically-relevant BESS scores, using the same detection criteria defined by the original BESS test. In order to assess the AAPS balance evaluation capability, 15 healthy subjects (7 male, 8 female) were required to perform the BESS test, while simultaneously being tracked by a Kinect 2.0 sensor and a professional-grade motion capture system (Qualisys AB, Gothenburg, Sweden). High definition (HD) videos with BESS trials were scored off-line by three experienced observers for reference scores. AAPS performance was assessed by comparing the AAPS automated scores to those derived by three experienced observers.

The most commonly used clinical balance assessment tool following concussion is the Balance Error Scoring System (BESS). The BESS test measures static postural stability and it is typically administered by trained medical personnel who must observe and count on a 0-10 scale, specific behaviors corresponding to deficits in postural control while simultaneously spotting the subject to prevent falls. The subject under test is required to maintain balance with eyes closed and hands on hips in three stance conditions: double-leg, single-leg and tandem stance. Each stance is performed on two surface types, hard ground (DS, SS, TS) and on a foam pad (DF, SF, TS). The standardized BESS defines the subject's balance errors, which must be counted:

- Moving the hands off the hips
- Opening the eyes
- Step, stumble or fall
- Abduction or flexion of the hip beyond 30°
- Lifting the forefoot or heel off the testing surface
- Remaining out of the proper testing position for longer than 5 seconds

Although fast and inexpensive, the BESS test presents a series of limitations that are intrinsically related to its subjective and manual scoring method. The BESS has been reported to have modest and widely ranging test sensitivity due to scoring inaccuracies and observer bias. It has been reported that the inter-rater and intra-rater minimum detectible change for the total BESS score were respectively 9.4 and 7.3 points. These changes are in the same range as BESS score differences between baseline and testing in concussed subjects. It is known from the literature that the average BESS score after concussion is 17 errors (range, 15-19 errors), compared with 10 errors at baseline (range, 8.4-12.7 errors). Further BESS limitations are the need for properly trained medical personnel to administer the test and its susceptibility to fatigue and practice effects.

The balance error detection algorithm has been implemented in the AAPS system to evaluate postural stability and provide a reliable and automated BESS score starting from raw Kinect sensor data. The algorithm has been designed to track balance errors as they are defined in the BESS standard. Subsequently, the AAPS extracts human body joint coordinates and locates the floor plane in real-time. The floor plane is used to identify the position and tilt of the sensor with respect to the subject. The joint coordinates are multiplied by a rotation matrix to compensate for sensor tilt and positioning. Next, the data frame rate is set to a constant value of 30 frames per second using linear interpolation. This is necessary because the Kinect provides data at a variable frame rate that depends on the instantaneous operating conditions of the acquisition computer (hardware/software) and data collection environment conditions such as lighting, room geometry, type and number of objects in the sensor field of view. To further account for the potentially large variability in the Kinect sensor frame rate (5-30 fps), the AAPS software was designed to perform real-time frame rate checks. If during a trial, the instantaneous frame rate drops below a certain value (10 fps in this application), an error message is displayed and the user is notified that the acquisition needs to be repeated. This is a fundamental feature in an automated system to guarantee acceptable performance levels in any condition. Based on our data collection sessions with the AAPS system, the ideal value of 30 fps tends to drop to 15 fps a few times per minute, while lower values are less frequent and usually occur once every 50 trials.

Next the extracted body joint 3D coordinates are filtered using a Savitzky-Golay filter. This is a smoothing filter with minimal signal distortion that operates by fitting low-order polynomial approximations to consecutive signal time windows using a least-squares approach. A filter with a third order polynomial approximation and a time window duration of 0.166 seconds was used. At a constant sampling frequency of 30Hz, such a window length corresponds to selecting five data points for each step of the least-squares approximation.

As discussed above, filtering Kinect data with a smoothing filter is necessary to attenuate the effects of the Kinect inaccuracy and variability in estimating the joint positions of a tracked human body, even when subjects stand perfectly still in the sensor field of view. With the signal adequately smoothed, body metrics are calculated on a frame-by frame basis. The metrics that have been used in the AAPS algorithm to detect balance errors during BESS trials are listed in Table 1.

Table 1: Calculated metrics extracted from Kinect raw data that are tracked during BESS tests

JOINTS OF INTEREST	METRIC $[M]$	DETECTED BALANCE ERROR $[E_i]$
Left hand – left hip	3D Distance	Hands off hips
Right hand – right hip	3D Distance	Hands off hips
Left Elbow – Left Hip	3D Distance	Hands off hips
Right Elbow – Right Hip	3D Distance	Hands off hips
Left Knee – Right Knee	3D Distance	Foot movement
Left Hip – Left Ankle	3D Distance (Single-Leg Stance)	Hip flexion
Right Hip – Right Ankle	3D Distance (Single-Leg Stance)	Hip flexion
Ankles	3D Position (Tandem Stance & Single-Leg Stance)	Foot movement
Frontal Plane Spine Angle	Angle	Spine frontal motion
Sagittal Plane Spine Angle	Angle	Spine sagittal motion

In order to detect errors in the subject's pose during balance trials, the algorithm uses a one-second calibration window to estimate the reference subject's stance and the current levels of noise in the Microsoft skeletal tracking algorithm. The calibration is necessary to assess data variability due to changes in both subject-specific poses and sensor-specific body estimations. Subsequently, the metrics are bandpass filtered (using a second order Butterworth filter) between 0.15Hz and 3Hz to emphasize signal components that are related to subject motion and to minimize other sources of variability (noise).

Additionally, sensor tracking inaccuracy is estimated by measuring the standard deviation of the noise in the calibration window. Specifically, the raw metrics are band-pass filtered with a second order band-pass Butterworth filter with passband set to 5-15Hz. This frequency range was selected to emphasize the signal components that are mainly due to measurement noise.

During the 20-second long BESS trials, the estimated calibration stance and the current subject's position are continuously compared. The comparison is carried out using a threshold that is set using the estimated standard deviation of the noise and the mean of the metric obtained during calibration. Balance errors are flagged each time the metrics cross such a threshold. Specifically, in the i-th frame, a balance error E_i is detected if the absolute difference between the calibration metric M_{cal} and the current metric M^i exceeds the threshold, set to ϵ times the estimated standard deviation σ_{cal} of the noise. The list of the kinematic metrics (M) that have been used to calculate the respective BESS errors (E_i) is presented in Table 1.

Mathematically, three categories of balance errors are detected:

1) Unilateral single threshold errors: estimated from low-noise and unilateral metrics.

$$E_i = \left(\left| \hat{M}_{n_i} - M_{n_{cal}} \right| > \epsilon_n * \sigma_{n_{cal}} \right)$$

where the subscripts *i*, *n* and *cal* indicate respectively the frame number, the type of metric and the calibration window.

2) Bilateral errors: estimated from low-noise bilateral metrics. An error is detected if the threshold is crossed on either side of the body.

$$E_i = \begin{array}{c} \left(\left|\widehat{M}_{n_{left_i}} - M_{n_{left_{cal}}}\right| > \epsilon_n * \sigma_{n_{left_{cal}}}\right) \\ OR \\ \left(\left|\widehat{M}_{n_{right_i}} - M_{n_{right_{cal}}}\right| > \epsilon_n * \sigma_{n_{right_{cal}}}\right) \end{array}$$

where the subscripts left and right indicate from which side of the body the metrics were derived.

3) Double threshold errors: to improve detection performance, errors, that are estimated using low-accuracy metrics, are detected using two correlated metrics and corresponding thresholds. An error is detected only if both metrics cross the threshold.

$$E_i = \left(\left|\widehat{M}_{n_i} - M_{n_{cal}}\right| > \epsilon_n * \sigma_{n_{cal}}\right) AND \left(\left|\widehat{M}_{m_i} - \widehat{M}_{m_{cal}}\right| > \epsilon_m * \sigma_{m_{cal}}\right)$$

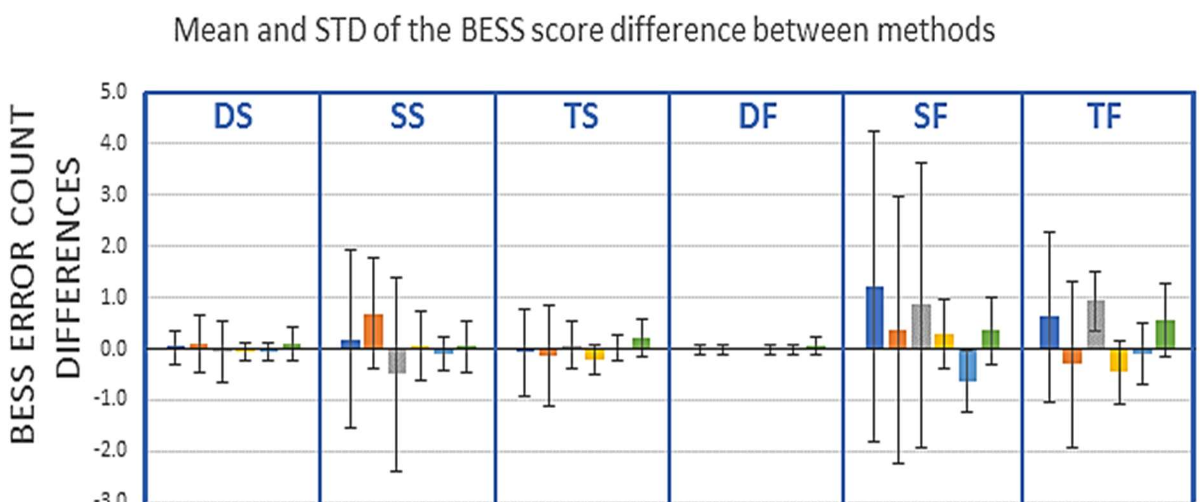
where the subscripts *n* and *m* indicate different metrics.

The above error types can be combined for improved balance detection precision. The different error types detected on a frame-by-frame basis are then converted into BESS scores, namely the total error count per trial, with two important caveats: 1) at most one error type can be detected within a pre-defined time window (set to 2 seconds); 2) a BESS error is recorded only if the infraction remains above the threshold for a pre-defined time duration (set to 110 ms).

In order to validate the results of the error detection algorithm, we simultaneously collected data using a Kinect sensor and a 12-Camera Qualisys system. Qualisys data have been post-processed using Opensim with a modified plug-in-gait model. After running inverse kinematics on the trajectory data, three-dimensional body joint positions were derived. The Kinect and Qualisys derived joint coordinate time series were time-synchronized using a large movement performed at the beginning of each trial and then fed into the BESS error detection algorithm as described above. Finally, scores obtained from the two systems were compared against scores from three human experts reviewing video footage of the BESS tests.

The AAPS algorithm was tested, using data derived from both Qualisys and Kinect systems, on 15 healthy subjects, each performing the BESS test twice. These subjects' balance was also evaluated by three expert observers using the gold standard BESS method. In the algorithm performance analysis, the average human scores have been chosen as ground truth for the correct error count.

Figure 1 shows the differences in the scores obtained using the different evaluation techniques: AAPS vs. Reference, Qualisys vs. Reference, AAPS vs. Qualisys, Human 1 vs. Reference, Human 2 vs. Reference, and Human 3 vs. Reference. The comparison of AAPS versus Qualisys was carried out to investigate potential differences in performance due to the two different optical acquisition systems. Variations in scores have been quantified by calculating the signed average difference between each technique and the reference. Differences can range between -10 and 10 points, where low error levels are indicated by values close to zero. Standard deviations are presented as error bars.



	DS	SS	TS	DF	SF	TF
AAPS vs. Ref	0.0	0.2	-0.1	0.0	1.2	0.6
Qual vs. Ref	0.1	0.7	-0.1	0.0	0.4	-0.3
AAPS vs. Qual	-0.1	-0.5	0.1	0.0	0.9	0.9
H1 vs. Ref	0.0	0.0	-0.2	0.0	0.3	-0.5
H2 vs. Ref	0.0	-0.1	0.0	0.0	-0.6	-0.1
H3 vs. Ref	0.1	0.0	0.2	0.0	0.4	0.5

Figure 1: Means and standard deviations of the score differences calculated for each balance scoring method and grouped by stance condition. Bottom: Mean error values for each group and condition. The tested stance conditions are: double leg (DS), single leg (SS) and tandem stance (TS) on firm ground; double leg (DF), single leg (SF) and tandem stance (TF) on foam pad. The blue, orange, grey, yellow, light blue and green bars represent different balance evaluations derived respectively for AAPS vs. Reference, Qualisys vs. Reference, AAPS vs. Qualisys, Human 1 vs. Reference, Human 2 vs. Reference, and Human 3 vs. Reference.

Table 2: Average differences expressed as percentage of agreement between different balance evaluation systems in detecting BESS scores, grouped by condition.

Percentage of Agreement						
Condition	AAPS vs. Ref	Qual vs. Ref	AAPS vs. Qual	H1 vs. Ref	H2 vs. Ref	H3 vs. Ref
DS	99.8	99.0	99.3	99.5	99.5	99.0
SS	98.1	93.1	95.0	99.5	99.0	99.5
TS	99.3	98.6	99.3	97.9	100.0	97.9
DF	99.8	99.8	100.0	99.8	99.8	99.5
SF	87.9	96.4	91.4	97.1	93.6	96.4
TF	93.8	96.9	90.7	95.5	99.0	94.5

Table 2 reports the overall level of agreement for the different groups, where values close to 100% (high agreement) correspond to differences in BESS scores close to zero. The values in the table are calculated by taking the percentage complement of the normalized absolute average differences in the scores. The absolute differences were normalized using the BESS full scale (10 points per trial).

To evaluate the statistical significance of the observed score variations a multiple comparison one-way ANOVA test was implemented (alpha set to 0.05). The results are shown in **Error! Reference source not found.Error! Reference source not found.**, where the means (filled circles) and 95% confidence intervals (horizontal lines) of condition-based balance scores are presented. The gray vertical dotted lines represent the 95% confidence intervals with respect to the Reference group. No statistically significant differences were found between any of the balance scoring methods and the Reference (average human scores, in blue). The multiple comparison ANOVA results emphasize that although differences in the scores are non-significant, the Kinect-based AAPS reaches its lowest performance in the single-leg on foam condition, as also highlighted by the lowest agreement levels reported in Table 2.

Although single-leg on foam was identified as the condition with lowest agreement levels between observers, there was no significant difference in performance between the Kinect-based and the Qualisys-based AAPS. This finding suggests that the AAPS software algorithm provides satisfactory performance levels using raw data from both motion capture systems; BESS error detection performance is not significantly affected by the acquisition hardware.

The ANOVA analysis did not reveal any significant difference in the scores. It is worth noting that the lowest AAPS performance levels are detected in single-leg and tandem stances on foam. In such conditions, despite the Qualisys-based AAPS system performing more closely to humans than the Kinect-based one, statistical analysis shows no significant difference in performance. This result demonstrates that the AAPS, built around an inexpensive, general-purpose 3D single-camera sensor, is viable for use in on-field applications.

The lowest agreements between both the AAPS systems and human observers are seen in the single-leg and tandem stances on foam condition. We hypothesize that lower agreement levels might be due to the higher levels of subjective evaluation that this condition requires to detect BESS errors. Specifically, we

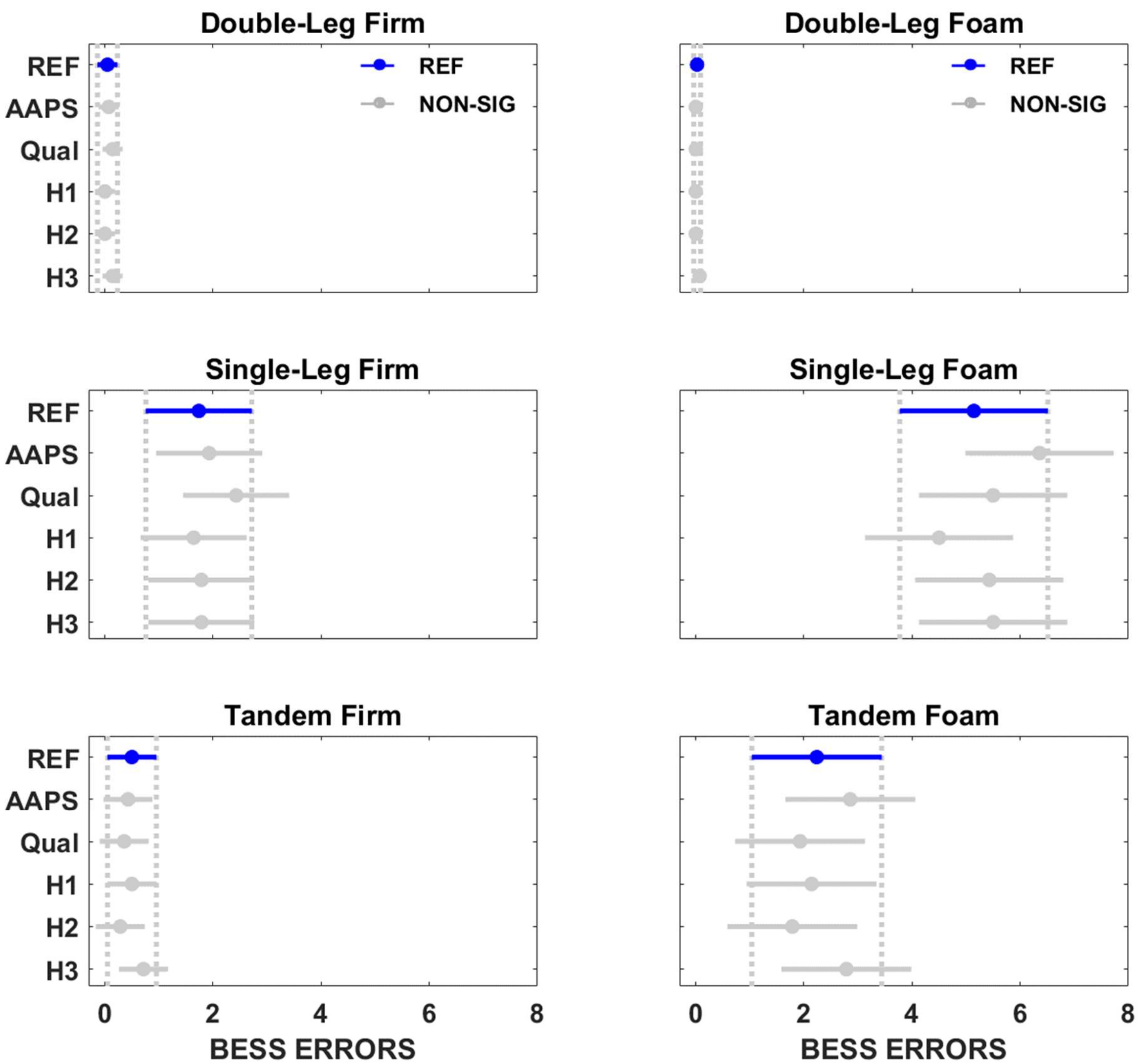


Figure 2: Results of a multiple comparison ANOVA test on the BESS scores. BESS errors derived using the AAPS, the Qualisys and three different human observers are compared to the average human scores, used as reference. Reference groups are in blue; Vertical dotted lines are 95% confidence intervals for the Reference group. None of the differences with respect to Reference are statistically significant.

identified three main factors. First, the presence of the foam complicates balance evaluation, because the foot on which the subjects stand is partially obscured by the foam. Secondly, this condition is arguably the most challenging, and consequently more motion is expected. This results in multiple errors and subjects having more difficulty to find and maintain their balance when trying to go back into the right position. In these cases, we found that human observers tend to use their "judgment" to count errors rather than strictly relying upon the BESS rules for balance error count. Finally, in single-leg on foam conditions, the automatic system seems to be operating at the limits of agreement between humans and AAPS systems because of the low sensitivity of the BESS test. This limitation has been reported in previous studies in which the modest sensitivity of the BESS is explained by the large variance in performance during the stances on foam. Over 53% of the variance in errors can be attributed to the single-leg and tandem conditions on foam.

The BESS only focuses on static postural control tasks and lacks assessment of more dynamic postural tasks. Thus, the choice of filtering the kinematic metrics between 0.15Hz and 3Hz to emphasize relevant data was deemed appropriate. The Kinect, and consequently the AAPS capabilities will be tested at their operational limit when introducing dynamic testing with the aim of capturing "faster" human movements. In such conditions, although the motion of large human body segments rarely exceeds a few Hertz, the filter high cut-off frequency needs to be increased to avoid signal's distortion and artifacts. However, based on our preliminary data during dynamic trials, the AAPS seems to perform at acceptable levels when compared to the Qualisys lab-grade performance.

Testing only for static stability may not capture other important domains of balance, including dynamic or multi-task postural control aspects. It is worth noting that the AAPS capability of detecting balance deficits had to be reduced to a single error count number per trial for the purpose of the comparison presented here.

These limitations derive from human administration of such testing protocols, wherein some information (e.g. error type, time, and magnitude) must be sacrificed in order to accommodate the capacity of a human observer. We hypothesize that an improved automated balance test, in which dynamic conditions and more reliable proxy kinematic variables are used, can be readily implemented by exploiting the existing capabilities of the AAPS system. The use of such a system to detect, track and quantify balance deficits in the field will provide the opportunity to go beyond traditional balance testing protocols that only rely on human visual observations reported with manual annotations. This will facilitate more informed and data-driven clinical decision making in non-clinical settings.

Despite some level of disagreement between human and AAPS-generated scores, the use of an automated system yields important advantages over currently available human-based alternatives. A computer scoring system is by definition deterministic, meaning that it eliminates variability during repeated evaluations, the same criterion does not apply to human scoring. Moreover, the AAPS can record specific error types with extremely high temporal resolution, it can detect multiple error types on a frame-by-frame basis and record their time course progression. These features are not achievable by humans, who cannot keep track of all those variables with such a high time resolution. Together, these results suggest that computerized BESS calculation may provide more accurate and consistent measures of balance than those derived from human experts.

Our results show that the AAPS error detection algorithm presented here can accurately and precisely detect balance deficits with performance levels that are comparable to those of experienced medical personnel. Specifically, our results show agreement levels between the AAPS algorithm and the human average BESS scores ranging between 87.9% (single-leg on foam) and 99.8% (double-leg on firm ground). In addition, statistically significant differences were not detected by an ANOVA test with significance level set to 0.05. Moreover, significant performance deficits were not detected when the less expensive, portable and marker-less AAPS was compared to a lab-grade system, with agreement levels between the two different motion capture systems ranging between 90.7% (tandem on foam) and 100% (double-leg on foam). These results underscore the value of using the Kinect-based AAPS, which can be quickly deployed in the field and/or in outdoor settings with minimal set-up time.

In future work, we plan on expanding the AAPS with new features, such as introducing criteria to account for balance error characteristics and fine-grained evaluation of dynamic and static postural control strategies using kinematic variables rather than trying to capture complex motion performance with an arbitrary summary scale. Such a system will also implement functional dynamic protocols that can be customized to a specific subject and application. These new dynamic posture screening tools combined with the ability to derive real-time meaningful postural metrics will help us develop innovative automated tools for more effective and comprehensive on-field postural strategy assessment. Furthermore, the AAPS capabilities will be tested in clinical populations, such as individuals suffering from low-extremity injuries and concussion.

Accomplishment 2: AAPS Field Testing

AAPS field testing capability and limitations have been thoroughly evaluated. Specifically, the system's accuracy and precision in correctly measuring body segment lengths under field-relevant conditions related to lighting, clothing and footwear have been investigated. The results of this work will be presented at the next "IEEE Signal Processing in Medicine and Biology Symposium" (SPMB17). The work sought to quantify the accuracy of the AAPS by measuring the lengths of body segments under various "real-life" conditions and to identify the ambient conditions that provide optimal results.

To evaluate the performance of the AAPS system in tracking body segment lengths, kinematic data were collected from five different subjects, with each subject performing a modified version of the Balance Errors Scoring System (BESS) under eight different experimental ambient conditions. Subjects were required to perform four of the six standard BESS stances, namely, single-leg and tandem stances on either firm ground or medium density foam pad. Each stance was repeated three times for each experimental condition. Experimental conditions were defined by combinations of three independent binary variables: types of environment, apparel, and footwear (Table 3). The indoor environment was a spacious office with fluorescent lighting and carpeted floors, while the outdoor one was the entrance area to an academic building which was removed from direct sunlight and had a concrete surface. Shoe type and color were not controlled.

In total, each subject was required to perform 96 trials: 3 repetitions for each of four poses under each of the 8 conditions. Each individual trial started with the subject facing the sensor with their arms spread out perpendicularly and their feet spread out shoulder width (the "T–pose"). After the T-pose, the subject engaged in one of two stances, single-legged stance or Tandem. Single-legged stance consisted of standing on one leg with hands resting on hips, while tandem stance entails standing with one foot directly in front of the other with hands resting on hips. During each trial, the subject executed one of three predetermined motions, providing the study with tracking data that replicates balance errors as defined in the BESS test.

Table 3: Variable Definitions

Variable	State 1	State 2
Environment	Indoor	Outdoor
Apparel	Long sleeve/ long pants	Short sleeve/ shorts
Footwear	Shod	Barefoot

The three predetermined motions were defined as follows: foot touchdown, where the subject takes a lateral step to catch his or her balance; trunk lean, where the subject performs hip flexion or abduction greater than 30°; hand-off hip, where the subject is unable to keep his or her hands resting on their hip.

The system performance analysis begins with a comparison between the Kinect-measured body segment lengths during the T-pose and the clinically-derived ones. The comparison was carried out using data derived with respect to all the ambient conditions. Later, we focused on the role of variable ambient combinations to identify the optimal condition that yielded the most accurate measurements. Finally, we analyzed the impact of the BESS stance and surface type on the system's tracking performance under the previously identified the optimal condition.

The Kinect v2 provides raw 3D coordinates for 25 body joint centers for each camera frame. The first step of the data processing was to set the data sampling frequency to a constant value of 30fps using linear interpolation. This is necessary because the Kinect sensor returns data at a variable frame rate (between 5-30 fps) depending on the computer's instantaneous processor demands, and is not user controllable. Once interpolated, the body joint 3D coordinates were fed into a 6th order Butterworth low-pass filter with a cutoff frequency of 3 Hz. Once the data were filtered, the length of each segment was calculated by taking the Euclidian distance between two joints that define each body segment.

Each subject had their body segments hand-measured by a single investigator and these measurements were compared to the length of each segment in the T-pose as estimated by the Kinect sensor. Each segment was measured using the following proximal and distal landmarks: acromion process to lateral humeral epicondyle (humerus); radial head to radial styloid (ulna); greater trochanter to lateral femoral condyle (femur); palpated joint space to lateral malleolus (tibia). The percent error was calculated per segment and used as a baseline to estimate overall differences between Kinect-estimated body segment lengths (Table 4) and clinically-derived metrics (Table 5) and the corresponding percentage (Table 6).

The effects of ambient condition on Kinect-derived body segment lengths were quantified by grouping the measurements by condition and BESS stance and then normalizing the results. Each subject's segments were first sorted by condition, yielding eight sets of segment data. Then, each set was divided into five groups, the four BESS stances with the additional T-pose stance, resulting in 40 sets (8 conditions x 5 stances) where each set represents a unique combination of ambient variables. To compare results across subjects, different segment and experimental conditions, we introduced the Normalized Root Mean Squared Error (N-RMSE) with respect to each set of data. The N-RMSE was derived using the following equation where KS denotes Kinect measured segment and CM denotes clinically measured segment.

$$\boldsymbol{NRMSE} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(KS_i - CM)^2}}{CM}$$

This metric was computed for each segment and condition type as reported in the Results section. The N-RMSE is a concise and unitless measure of error that allows for an easy comparison between several variables and varying experimental conditions. Finally, the N-RMSE's were averaged across subjects, to emphasize error trends mostly dependent on the tested conditions and stances that were common to all the subjects.

In order to carry out a fair comparison between clinically-derived metrics and Kinect-derived ones, we utilized normalized differences. The Kinect's joint tracking algorithm is proprietary, thus it is not feasible to know exactly where on the body the joint centers are placed. In other words, we expected the Kinect-measured lengths to be different from our clinically-derived ones because of the different techniques, and therefore we compared the relative change across different ambient conditions rather than the absolute measurements. Tables 4-6 present the Kinect-derived measurements, the clinically-derived ones and the percentage errors of their differences averaged across all the ambient conditions. These values quantify the initial error between the two measurement techniques without accounting for any ambient condition changes. We found an average measurement error of 10.8%.

Table 4: Kinect v2 T-pose Measurements. These represent the average segment measurements averaged across the ambient conditions

Subject	Tibia (cm)	Femur (cm)	Ulna (cm)	Humerus (cm)
1	42.5	32.1	22.9	24.3
2	44.7	41.5	23.6	26.2
3	42.5	36.2	23.4	25.9
4	45.5	38.9	23.5	24.7
5	43.2	37.6	23.5	24.7

Table 5: Clinically-derived measurement of each subject's segment measured by a single investigator.

Subject	Tibia (cm)	Femur (cm)	Ulna (cm)	Humerus (cm)
1	40.1	38.1	26.7	25.4
2	40.4	45.7	23.5	29.2
3	40.0	40.6	26.7	28.4
4	41.1	42.4	26.2	26.9
5	42.5	41.5	25.5	31.0

Table 6: Percent Error. The percentage error difference between clinically-derived (Table 4) and Kinect-derived lengths (Table 5).

Subject	Tibia (%)	Femur (%)	Ulna (%)	Humerus (%)
1	8.3	16.5	14.2	4.8
2	11.4	10.5	4.4	10.4
3	9.2	12.9	12.6	8.9
4	4.9	10.5	7.6	21.9
5	8.8	8.6	9.2	20.3

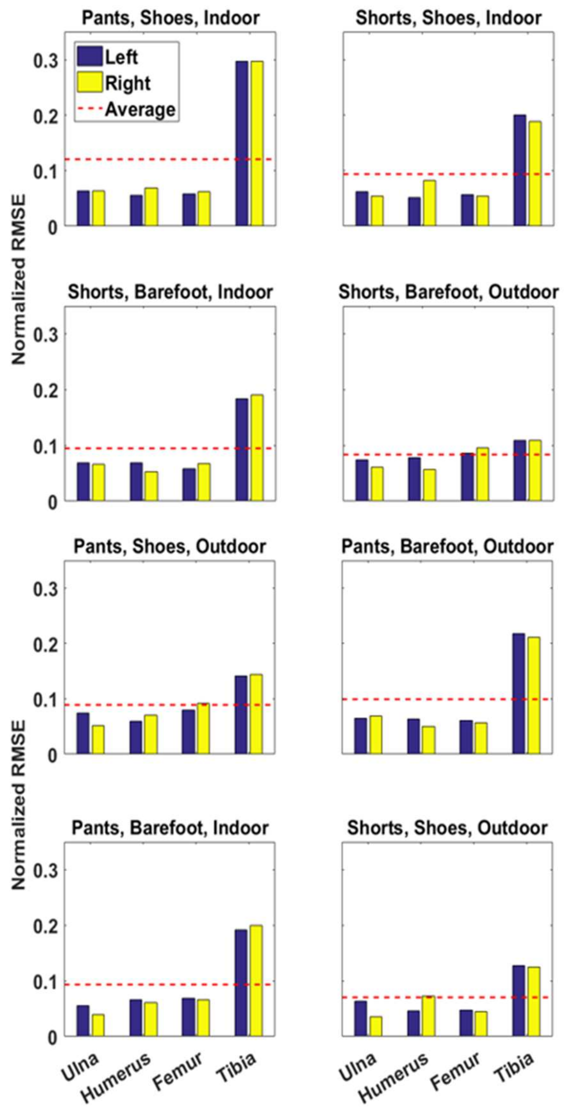


Figure 3: The normalized RMSE for each segment in each condition of the T-pose stance. Blue and yellow bars represent body segment N-RMSE's for left and right side of the body, respectively. The dashed red line represents the overall N-RMSE averaged across all segments.

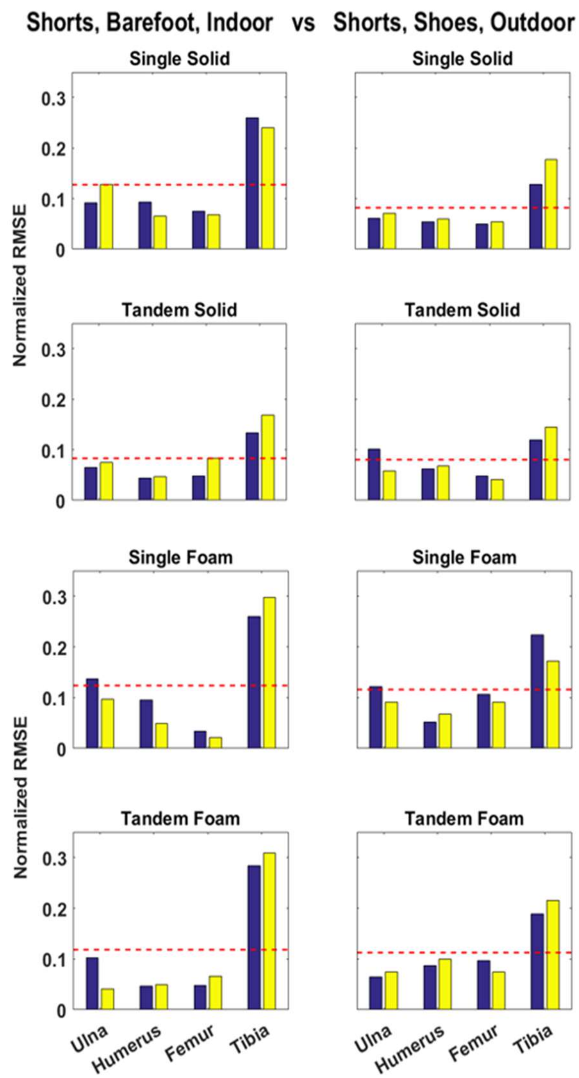


Figure 4: The normalized RMSE for each segment in each stance and surface of the hypothesized ideal condition and the actual ideal condition which are the left and right columns, respectively. Blue and yellow bars represent body segment N-RMSE's respectively for left and right side of the body. The dashed red line represents the overall N-RMSE averaged across all segments.

Subsequently, we introduced the N-RMSE to quantify the amount of error for each segment in each condition and stance. Given the large number of conditions and trials, we initially focused only on segment lengths in the T-pose stances in order to identify optimal ambient conditions. The T-pose stance was selected because it had the least amount of motion and allowed for optimal limb visibility, meaning that errors would mostly be caused by the ambient conditions. Analysis results are shown in Figure 3. Although we hypothesized that the optimal ambient condition could have been a trial with the subject wearing shorts, no footwear and indoor, the data indicated that the condition with the lowest error levels was when the subject wore shorts and shoes outdoors, as indicated by the lowest overall average N-RMSE for this condition, reported with a red dashed line in Figure 3

We next investigated the effects of stance and surface on the Kinect's accuracy. Two ambient conditions were included in this analysis: 1) shorts and shoes, outdoor (the observed optimal condition) and 2) shorts, no shoes, indoor (the hypothesized optimal conditions). Results are shown in Figure 3, where the former condition outperforms (lower errors) the latter in every stance. This figure also emphasizes how lower extremities are negatively affected by the presence of the foam pad, especially in the tandem stance.

Figure 4 shows the N-RMSE across ambient conditions in the T-pose. It is worth noting how the conditions with shoes are typically better overall than the ones without. This is due to the Kinect's depth camera that can detect a shoe more easily than a bare foot because of the shoe's higher contrast with the ground.

Figure 4 shows the N-RMSE across stances in both the hypothesized ideal condition (on the left) and the actual ideal condition (on the right). In Figure 4, it is also of interest to notice how stances with foam have larger error than stances on firm ground. This is due the Kinect sensor estimating the floor plane without taking the elevated foam pad into account. This results in elongated leg measurements.

This work evaluated the accuracy of various Kinect-derived body segment length measurements as a function of clothing, environment, and footwear. Within these conditions, subjects were tested in stance tasks based on a subset of those used in the BESS test. Our analyses revealed that normalized RMSE (based on a clinically-obtained reference measurement) was greatest for the tibia across experimental conditions and across stance tasks. Error levels for the remaining segment lengths varied little based on clothing/environment/footwear. Although we had initially hypothesized that segment length estimation would be most accurate in the indoor, barefoot, shorts condition, our results suggest that the optimal combination of the tested conditions, based on normalized RMSE, may be outdoor, shoes, shorts. Both of these conditions were analyzed in greater detail to determine the influence of stance task. No effect of stance task was observed for segment length measurement of the upper extremities. In both conditions, standing on foam adversely impacted the Kinect's segment length estimation performance at the tibia. The average N-RMSE in tibia segments was 0.1714 on solid ground and 0.2440 with foam, an increase of 42.36%. These same patterns were not observed as clearly when estimating the femur. In shorts, barefoot, indoor, standing on foam was associated with lower normalized RMSE at the femur. The opposite relationship was observed in the shorts, shoes, outdoor condition. The results of this study suggest that Kinect-based segment length estimation is optimized in outdoor environments with the subject wearing shoes and shorts. Segment length estimation is least accurate for the tibia, which is also the segment most susceptible to adverse effects associated with medium density foam commonly used in clinical balance testing.

Expanded AAPS for Dynamic Motion

The expanded version of the AAPS, the xAAPS has been developed and a beta software version has been implemented. The new system has been tested on a total of 30 subjects. The xAAPS is a postural stability system that evaluates dynamic balance tasks. By testing the ability to make coordinated dynamic movements and maintain balance, xAAPS system can potentially provide more salient feedback for assessing suitability for return to duty than using static balance measures alone. The battery of dynamic tests implemented in the xAAPS for postural assessment includes:

- sit-to-stand
- hurdle step
- deep squat
- in-line lunge
- time-to-stabilization
- marching in place

The xAAPS system is currently capable of running dynamic trial data acquisition and it has been designed to provide real-time feedback to guide subjects during the execution of trials. Furthermore, the system will provide real-time metrics on the motion quality. Our team is currently working on developing optimal algorithms to reliably evaluate and quantify balance and motion strategies, for both online and off-line kinematic data analysis. The goal is to translate into deterministic and data-driven criteria some of the commonly used movement evaluation guidelines. The xAAPS will be able to capture and automatically generate data-driven scores that correlate with the “gold-standard” manual assessments.

The new algorithms are being prototyped in Matlab, using previously collected data. Once such algorithms reach optimal performance levels, they will be ported to C# code and introduced into the xAAPS code infrastructure as part of the new and expanded postural stability suite.

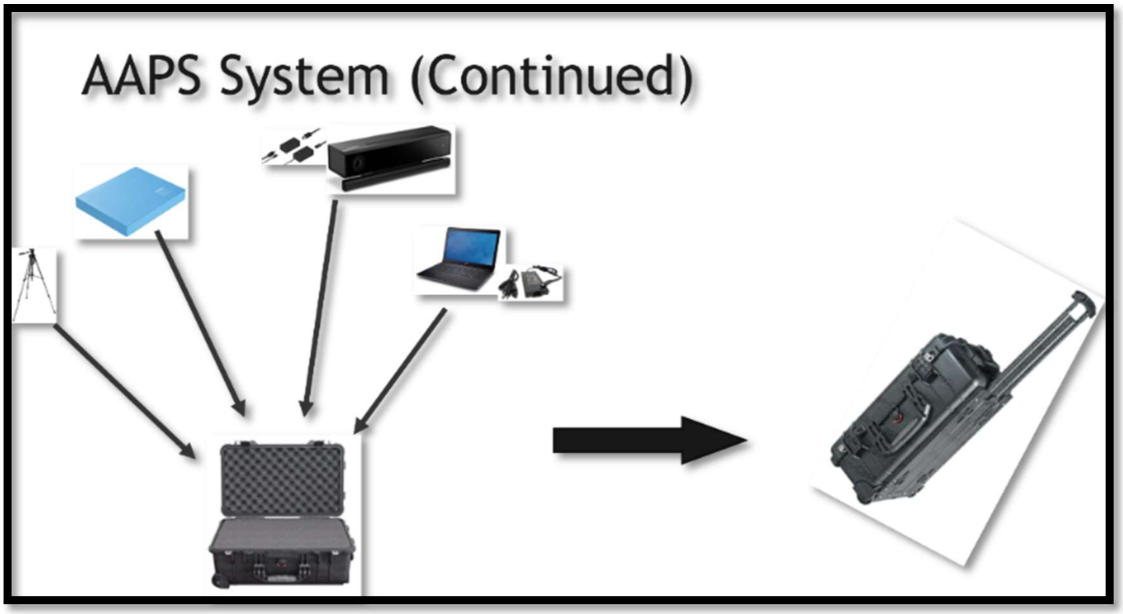
Training & Training Materials

Training materials and system documentation were created to ensure maximum operability by any user. These materials include 1) a technical user manual detailing all aspects of operating the hardware and software, and 2) slide-based training modules to minimize time from “out-of-the-box” to “up-and-running.” We have thus far trained several non-clinician users to administer AAPS testing successfully. Moving forward, we will conduct a structured training and feedback cycle to ensure that use of the system by new parties is maximally streamlined. Figure 5 shows selected figures from the training materials.

Field Testing

We have extensively field-tested the AAPS system. Our first endeavor in this regard was to determine and document the design features required for military-grade ruggedization of the AAPS system. This was undertaken using input from our military advisory panel and additional investigation into the hardware/software limitations of the AAPS system’s components. Our considerations for environmental ruggedization, as well as maximizing usability within the constraints of military, are included the system’s documentation.

Additional field-testing was conducted on an outdoor BMX bicycle course. This venue provided a lightly wooded area with hilly terrain and dirt/clay trails. Based on previous meetings with our military advisory panel, we deemed that this would reasonably replicate the most concerning environmental (i.e. non battle-related) challenges associated with in-theatre field use. Preliminary analyses indicate that body tracking and error detection perform well despite the constrained environment, variable lighting and stance surfaces, background clutter, and background motion.



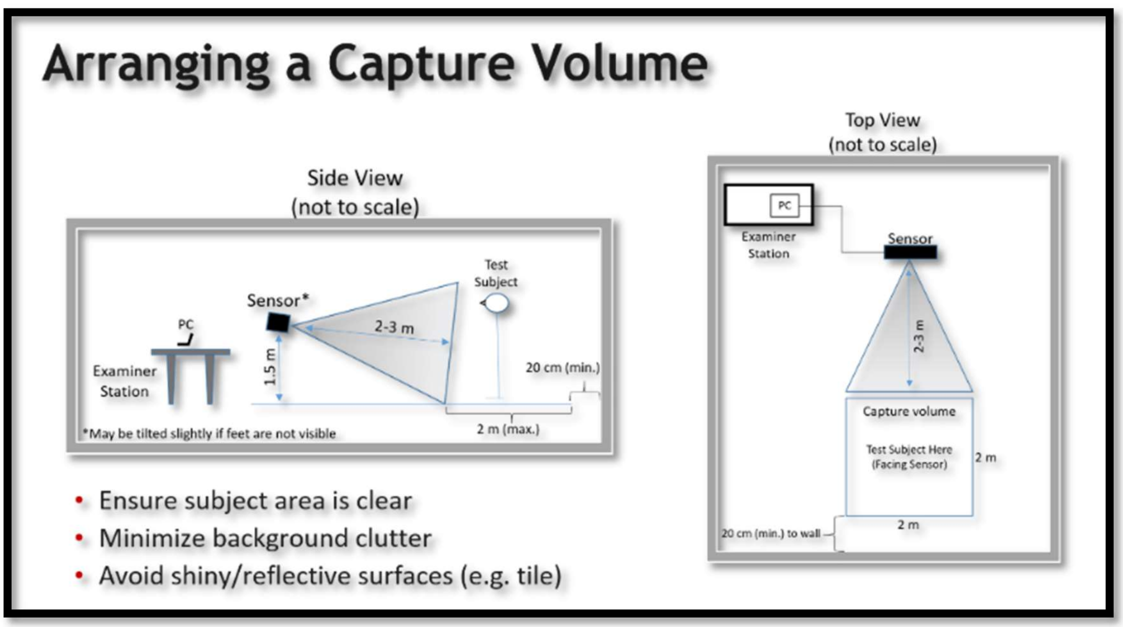


Figure 5: Selected images from training materials

<u>Progress Relative to Goals</u>

Relative to our stated goals, the project is proceeding on schedule and under budget. As Table 7 shows, our progress is largely commensurate with the 24 months of effort we have made thus far. Aim 1 is essentially complete, save for a few minor outstanding details. Under Aim 2, we have met our recruitment goals for healthy subjects and continue to seek new ways of recruiting concussed or injured individuals. Under Aim 3, we have evaluated the AAPS under a number of non-laboratory conditions with respect to lighting, background clutter, and subject clothing. We have developed a thorough training module for non-clinician operators which we will be evaluating and improving over the coming year. Finally, under Aim 4, we have made significant progress in upgrading the AAPS to handle dynamic movements instead

Table 7: Project status relative to timeline originally stated in the research proposal.

Specific Aim 1 – Develop AAPS Baseline System		
Port Image Processing Code to C/C++	1-5 months	100%
Develop User Interface	4-8 months	100%
Develop AAPS for Field Use	7-12 months	90%
Specific Aim 2 – AAPS Calibration and Baseline Evaluation		
Healthy Subject Evaluation	12-18 months	75%
Concussion Subject Evaluation	18-30 months	20%
Mild Musculoskeletal Injury Subject Evaluation	18-30 moths	31%
Specific Aim 3 – AAPS Field Evaluation		
Evaluate use by non-clinician operators	12-15 months	50%
Evaluate AAPS in Field Conditions	14-24 months	75%
Specific Aim 4 – Develop Expanded xAAPS Test		
Determine movements for xAAPS test	18-22	50%
Update AAPS software for xAAPS test	18-30	50%
Evaluate xAAPS test	30-36	10%

of just static balance poses. The software for motion tracking is largely complete and we will continue to harden and test it over the coming year.

From a budgetary perspective, the project is healthy. As of the end of Year 2 Quarter 4, we have spent $821k, which is about 10% under budget. The expense breakdown is approximately 59% compensation expenses, 6% non-compensation expenses, and 35% indirect costs.

What opportunities for training and professional development did the project provide?

This project has provided excellent opportunities for training and professional development, since almost all of the main work has been performed by trainees. Although the co-PIs retain close oversight, day to day operations and planning have been delegated to the postdoctoral fellows, for whom this is excellent professional training. The graduate research assistants have been mentored by both PI Obeid and the Fellows. Their responsibilities have included writing software and developing much of the back end mathematics behind the manipulations of the three dimensional mathematics. The four undergraduates employed this year (three female) have learned to program, to debug software, and to work use team-based software tools such as bug tracking and distributed version control. They have contributed meaningfully to the development of this work. Both postdoctoral fellows have attended research conferences this year and all members contributed to planning and writing the manuscripts.

How were the results disseminated to communities of interest?

This year, the team is pleased to report that we have published two journal articles with a third currently in review and a fourth under development, as well as a conference publication. On a more informal level the team has had a number of interactions with outside teams during which technical expertise and findings have been communicated. The Military Health System Research Symposium continues to be an excellent venue for such interactions.

What do you plan to do during the next reporting period to accomplish the goals and objectives?

During Year 3, we expect the emphasis to be on the xAAPS and completing our data collection and analysis. We are on track to complete all stated tasks by the end of Year 3.

Impact

Principal project discipline

Although there have been a handful of reports in the literature describing how accurate the Kinect is, none of them have adequately taken into account either the range of 'normal' movements (in a kinematic sense) or the allocation of errors into the three cardinal planes. Such data is critical in order to understand the limits of accuracy that can be expected from Kinect-based systems. By publishing our findings in this area, we are making a fundamental contribution to the Kinect-based motion tracking research community. Furthermore, we are demonstrating how Kinect and other consumer grade motion tracking systems can be used to quantify more natural movements in more degrees of freedom than in any other published applications.

Other disciplines

Nothing to report.

Technology transfer

Nothing to report (yet).

Society Beyond Science and Technology

The goal of this project has always been to improve awareness and treatment of concussion by providing a low cost, low complexity way of quantifying degradation in balance ability. Given what is now known about the pervasiveness of concussion and mild brain injury, especially in the armed forces community, this project has the potential to contribute positively to societal health and wellbeing.

Changes/Problems

Changes in approach and reasons for change.

Nothing to report.

Actual or anticipated problems or delays and actions or plans to resolve them.

We have encountered some difficulty in recruiting subjects to meet our clinical population targets. We have pursued alternative sources of clinical subjects, particularly focusing on the TBI/mTBI population, to include mixed martial arts and CrossFit gyms, BMX cycle athletes, and area recreational sports leagues. We recently arranged to host a booth at the International Flat Track Derby Association championship tournament this November (3^{rd} – 5^{th}), during which we hope to have access to a number of concussion and lower extremity injury participants. We will continue to pursue all avenues and have begun preparations to modify our IRB to permit subject remuneration should our enrollment targets not be reached in the near future.

Changes that have a significant impact on expenditures.

Nothing to report.

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents.

Nothing to report.

Products

[1] Napoli, A., Ward, C. R., Glass, S. M., Tucker, C., & Obeid, I. (2016). Automated Assessment of Postural Stability System. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Vol. 2016, pp. 6090–6093). IEEE. http://doi.org/10.1109/EMBC.2016.7592118

[2] Napoli, A., Glass, S. M., Ward, C. R., Tucker, C., & Obeid, I. (2017). Performance Analysis of a Generalized Motion Capture System Using Microsoft Kinect 2.0. *Biomedical Signal Processing and Control*, *38*, 265–280. http://doi.org/10.1016/J.BSPC.2017.06.006

[3] Napoli, A., Glass, S. M., Tucker, C., & Obeid, I. (2017). The Automated Assessment of Postural Stability: Balance Detection Algorithm. *Annals of Biomedical Engineering*. http://doi.org/10.1007/s10439-017-1911-8

[4] Glass, S. M., Napoli, A., Obeid, I., & Tucker, C. (2017). Inverse Kinematics Using Portable, Low-Cost Sensor Technology. In *Proceedings of the Gait and Clinical Movement Analysis Society Annual Meeting*. Salt Lake City, UT.

[5] Glass, S. M., Napoli, A., Obeid, I., & Tucker, C. (2017). Equivalence Characteristics of Musculoskeletal Models Generated from Concurrently Acquired Markerless and Criterion-Referenced Motion Capture Systems. *Gait and Posture*, IN REVIEW.

[6] Glass, S. M., Napoli, A., Thompson, E., Obeid, I., & Tucker, C. (2017). Measurement Properties of an Automated Balance Error Scoring System. *IN PREPARATION*.

Participants & Other Collaborating Organizations

What individuals have worked on the project?

Name:	Iyad Obeid, PhD
Project Role:	co-Principal Investigator
Person-Months:	3
Contribution:	Dr. Obeid contributed to project design and management, analyzed data, supervised data marshalling, wrote quarterly reports, and contributed to all technical publications.

Name:	Carole Tucker, PhD
Project Role:	co-Principal Investigator
Person-Months:	3
Contribution:	Dr. Tucker contributed to project design and management, IRB preparation, human subject protocol design, data collection, and analysis, and all technical publications.

Name:	Alessandro Napoli
Project Role:	Postdoctoral Fellow
Person-Months:	12
Contribution:	Was responsible for managing all aspects of the software organization and development, and contributed heavily to actual software creation. He managed the graduate RAs and the undergraduates, contributed to data collection and analysis, and took a leading role on all technical publications.

Name:	Stephen Glass
Project Role:	Postdoctoral Fellow
Person-Months:	12
Contribution:	Was responsible for managing all aspects of data planning, collection and analysis, including IRB development. He managed junior students, and took a leading role in all technical publications.

Name:	Christian Ward
Project Role:	Graduate Research Assistant
Person-Months:	6
Contribution:	Provided software development and data analytics support; contributed to management of undergraduate students.

Name:	Anirvan Majumdar
Project Role:	Graduate Research Assistant
Person-Months:	2
Contribution:	Code development and documentation

Name:	Nicholas Satterthwaite
Project Role:	Graduate Researcher
Person-Months:	4.6
Contribution:	Code development and documentation

Name: Victor Espinoza
Project Role: Graduate Researcher
Person-Months: 3.5
Contribution: Code development and documentation

Name: Bhautik Amin
Project Role: Undergraduate Researcher
Person-Months: 1.2
Contribution: Code development and documentation

Name: Paula Oliveira
Project Role: Undergraduate Researcher
Person-Months: 0.4
Contribution: Code development and documentation

Name: Lillian Veloso
Project Role: Undergraduate Researcher
Person-Months: 0.9
Contribution: Code development and documentation

Name: Elizaveta Ibeme
Project Role: Undergraduate Researcher
Person-Months: 1.0
Contribution: Code development and documentation

Has there been a change in the other active support of the PD/PI(s) or senior/key personnel since the last reporting period?

Nothing to report.

What other organizations have been involved as partners?

Nothing to report.

Special Reporting Requirements

See Quad Chart in the Appendix

Appendices

Quad Chart – see next page

Automated Assessment of Postural Stability (AAPS)

Log Number: MR141272

Award Number: W81XWH-15-1-0445

PI: Iyad Obeid & Carole A. Tucker **Org:** Temple University **Award Amount:** $1.36M

Study/Product Aim(s)

•Develop a fully functional proof-of-concept system (AAPS), featuring a complete software suite for automatically administering the Balance Error Scoring System (BESS) test.

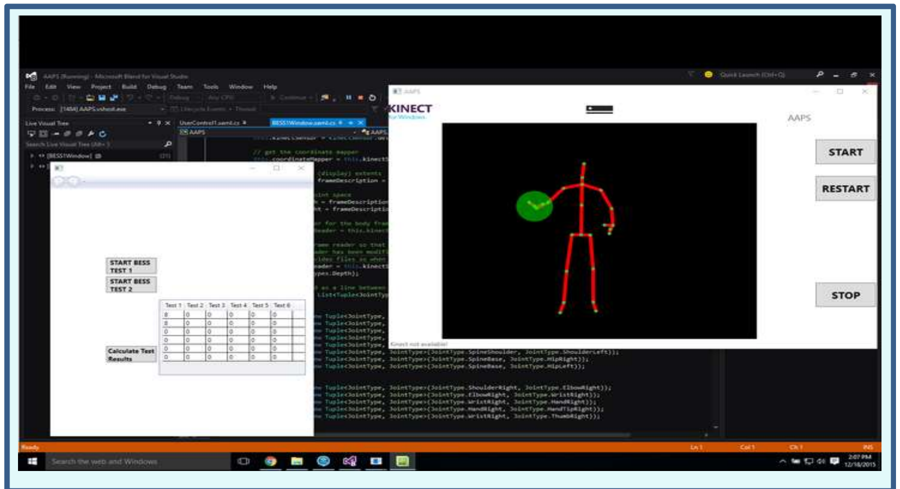
•Calibrate the AAPS on healthy, concussion, and musculoskeletal injury subjects.

•Fully field test AAPS to ensure use by non-medical technicians.

•Expansion of AAPS to include dynamic postural tasks.

Approach

We aim to develop, calibrate, and field test a system for quantifying the impact postural and balance injuries using the Microsoft Kinect, an inexpensive motion capture system. The system will administer and score the BESS in field conditions without requiring a medically trained operator. We will expand the BESS to include dynamic tasks (lunge, squat, etc.) to better assess readiness for return to active military duty post mild TBI.



Screenshot showing development of C# code, prototype proof-of-concept GUI interface, and live image capture skeleton.

Timeline and Cost

Activities **CY**	**15**	**16**	**17**	**18**
AAPS system development				
Calibrate AAPS (n=50 subjects)				
Field test AAPS				
Expand AAPS – Dynamic tasks				
Estimated Budget ($k)	**$200**	**$500**	**$500**	**$200**

Updated: 14 October 2017

Goals/Milestones

CY15 Goals – System development

✓ Port existing system from Matlab to C/C++ *[100%]*

✓ Develop user interface for automatic test administration *[100%]*

CY16 Goal – Calibration and Field Testing

✓ Determining reference scores for healthy, concussion, and musculoskeletal injury subjects *[55%]*

✓ Comparing performance to gold standard benchmarks *[85%]*

☐Optimizing design for use by non-medical technicians *[50%]*

CY17 Goal – System expansion

☐Determining optimal dynamic tasks for assessment *[50%]*

☐Updating software to handle dynamic task tracking *[50%]*

CY18 Goal – System optimization

☐Complete expansion & optimize software via beta testing *[10%]*

Comments/Challenges/Issues/Concerns

- none

Budget Expenditure to Date

Projected Expenditure: ~$907k

Actual Expenditure: ~$821k