Objective profiling of varied human motion based on normative assessment of magnetometer time series data

Claire M. Barnes[1], Cain C. T. Clark[2,3], Paul Rees[1], Gareth Stratton[2] and Huw D. Summers[1]

[1] Systems and Process Engineering Centre, College of Engineering, Swansea University, Bay Campus, Fabian Way, Swansea, SA1 8EN

[2] Engineering Behaviour Analytics in Sport and Exercise (EBASE) Research group, School of Sports and Exercise Sciences, Swansea University, Bay Campus, Fabian Way, Swansea, SA1 8EN, U.K.

[3] University Centre Hartpury, Higher Education Sport, Gloucestershire, GL19 3BE, U.K.

ABSTRACT

Objective: To quantify varied human motion and obtain an objective assessment of relative performance across a cohort. *Approach*: A wrist-worn magnetometer was used to record and quantify the complex motion patterns of 55 children aged 10-12 years old, generated during a fundamental movement skills programme. Sensor-based quantification of the physical activity used dynamic time warping of the magnetometer time series data for pairs of children. Pairwise comparison across the whole cohort produced a similarity matrix of all child to child correlations. Normative assessment scores were based on the Euclidean distance between *n* participants within an *n-1* multi-variate space, created from multi-dimensional scaling of the similarity matrix. The sensor-based scores were compared to the current standardised assessment which involves binary scoring of technique, outcome and time components by trained assessors. *Main Results*: Visualisation of the relative performance using the first three axes of the multi-dimensional matrix, shows a 'performance sphere' in which children sit on concentric shells of increasing radius as performance deteriorates. Good agreement between standard and sensor scores is found, with Spearman rank correlation coefficients of the overall activity score in the range of 0.62-0.71 for different cohorts and a kappa statistic of 0.34 for categorisation of all 55 children into lower, middle, upper tertile and top 5% bands. *Significance*: By using multi-dimensional analysis of similarity measures between participants rather than direct parameterisation of the physiological data, complex and varied patterns of physical motion can be quantified, allowing objective and robust profiling of relative function across participant groups.

Keywords: magnetometer, time series analysis, dynamic time warping, movement skill, physical activity, multi-dimensional scaling

1. BACKGROUND

The accurate measurement and recording of physical motion is an important aspect of many health and wellbeing programmes(Maud and Foster 2006, Penedo and Dahn 2005, Lynch *et al* 2014). Automated, objective measurement can be achieved with wearable, electronic sensor technologies(Brandes *et al* 2006, Godfrey *et al* 2008, Ermes *et al* 2008, Mathie *et al* 2004) and in the main these have focused on the quantity of physical activity or energy expenditure. In practise, physical activity measurement programmes can follow product-oriented analyses that require such quantitative measures of outcome, e.g. time taken or distance moved(Rudd *et al* 2016, Lander *et al* 2017), but alternatively they may be process-oriented and seek qualitative assessment of how a movement is performed. Thus, in many instances a more nuanced approach is required to ascertain the quality of movement, as for example, in the assessment of fundamental movement skills(Morgan *et al* 2013). This brings a measurement challenge as varied and complex motion patterns are inherently difficult to parameterise and as a result they are often judged by aesthetic criteria. Standardisation can be established through formal scoring schema but a tendency towards variability remains as approaches are fundamentally based on subjective human assessment(Barnett *et al* 2014).

In this paper, we present a normative assessment approach which is based on correlation of motion patterns, as represented in time series data from magnetometer sensors. The quantification process is based on the whole of the activity trace and scores the relative performance of individuals within a cohort. To demonstrate the validity of our approach we benchmark the sensor-based scoring against standard metrics for a children's activity programme, designed to assess fundamental movement skills(Stratton *et al* 2015). This application area is appropriate for whilst physical activity has long been recognised as being beneficial to children's health(Sallis *et al* 2000), reducing the risk of disease(Sirard and Pate 2001), combating obesity(Hills *et al* 2011) and improving non-health indicators such as educational achievement(Coe *et al* 2006); there is growing appreciation of approaches that provide greater discrimination in the assessment of activity. For example, techniques focusing on motor competence(Robinson *et al* 2015), fundamental movement skills (FMS) (Fisher *et al* 2005) and holistic concepts such as physical literacy(Tremblay and Lloyd 2010). The aim here is to assess the content or quality of human movement rather than just its intensity or duration(Myer *et al* 2015). This move towards more complex and subtle discrimination of movement introduces significant challenges to attaining formalised and quantitative metrics of outcome(Janz 2006, Trost 2001).

Inertial measurement devices have been used with signal analysis routines to machine-score specific activities or components within a varied activity programme (Bisi *et al* 2017, Allen *et al* 2006). This paper presents an alternative, process-oriented quantification of complex motion in which pairwise comparison of individuals is made using time trace correlations of position sensor data(Barnes *et al* 2016). We take a novel approach in analysis - instead of directly following the standard scoring protocol and trying to computationally identify specific motion parameters within each activity we extract comparative measures between children based on the whole of their activity trace. We then use the relative ranking of a study cohort to generate overall performance scores. Our premise is that during the performance of a varied sequence of activities the defining character in quality of movement or perfection of skills lies in the precise way in which a child moves through space. Therefore, if this motion pattern is captured by a sensor and then mathematically compared for different children, we can obtain a meaningful measure of their relative ability. Motion traces are recorded using a magnetometer attached to the participant's wrist as they progress through a series of nine discrete exercises. Previous approaches using wearable sensors tend to focus on identification of specific gestures(Akl *et al* 2011) or discrimination between specific activities, e.g. walking or cycling(Mannini *et al* 2013). Here we consider the whole complex motion pattern and analyse the complete exercise sequence rather than deconstructing the activity set. We measure the similarity between any two chosen children using Dynamic time warping (DTW) – a time series analysis technique which provides cross-correlated differences between the data sets, at multiple time points(Zhou and De La Torre 2016). DTW has been shown to be of use for human motion analysis, enabling classification of physical motion motifs(Ten Holt *et al* 2007, Musculo *et al* 2007). The full set of pairwise comparative coefficients from DTW allows us to construct a similarity matrix for the whole group. This matrix contains information on all possible relationships between the measured cohort. To visualise these and to obtain a unique ranking of individuals we adopt a data clustering approach(Braun *et al* 2010) and use Multidimensional Scaling algorithms(Borg and Groenen 2005, Hout *et al* 2013) to transform the similarity matrix into a map of relative performance.

Our aim is to demonstrate the general validity of the motion-based, normative approach. We do not attempt to undertake a detailed comparative study of specific clinical application or exercise programme. The analysis techniques are demonstrated on data from a movement skills programme for children. But they are applicable for any motion study that requires assessment of overall movement pattern rather than limited sets of biomechanical metrics and which seeks to rank or compare subjects. For example, in profiling gait abnormality or deterioration against a healthy population(Hausdorff *et al* 1998, Halliday *et al* 1998) or in quantifying return to full physiological competence following sports injuries(Andrew *et al* 2010, Podlog and Eklund 2006).

2. METHODS

The motion tracking sensor

The motion sensors used were custom built Micro Electro-Mechanical System (MEMS) based devices, that have been validated in previous studies(Barnes *et al* 2016). The sensors incorporated a tri-axial accelerometer with a +/- 16 g dynamic range, 3.9 mg resolution (ADXL345 sensor, Analog Devices) and a magnetometer with a +/- 8 Gauss range, 2 milli-Gauss resolution (HMC5883L sensor, Honeywell). The devices were set to record to an on-board micro-SD memory card at an acquisition frequency of 40 Hz. All data analysis was carried out using the MATLAB 2015b programming environment (Mathworks, U.S.A.) using in-built or custom-written functions.

Participants and settings

After receiving clearance to implement the study from the institutional research ethics committee, 55 children (33 female, 22 male) from 4 schools (school A – 20, B – 14, C – 12, D - 9) (11±0.5y, 1.45±0.06m, 40.4±9.4kg, body mass index; 19±3.5 kg.m^2) agreed to take part in the study. Participants had anthropometric variables recorded using standard techniques(Lohman *et al* 1988). Participants attended an indoor training facility where they took part in a multi-component assessment of physical activity. During this, participants wore motion sensors housed in a small plastic case and affixed via a Velcro strap to their right ankle (lateral malleolar prominence of the fibula) and dominant wrist (posterior wrist joint). All children from a single school were assessed in a single session, each school was assessed in a separate session on differing days.

The activity assessment was undertaken as part of the Sport Wales Dragon Challenge(Stratton *et al* 2015). Dragon Challenge V1.0 (DC V1.0), is a single practical assessment and designed to measure the stability, locomotor and object control components of fundamental movement skills of children in school years 6 to 7 (10-12 years old). In the DC V1.0 children complete 9 activities in a continuous circuit in a timed trial – there are no waiting periods, each activity is performed immediately following the previous one. Participation in the challenge requires spatial awareness (changes in direction and levels) and awareness of effort (changes in speed, force and flow) in relation to various objects, goals, and boundaries. Participants also utilise important cognitive attributes such as confidence, decision-making and reading the environment as they navigate through the tasks against the clock. A schematic, depicting each of the activities can be found in the DC V1.0 manual(Stratton *et al* 2015), in sequence they are:

1. *Balance Bench*: Walk the length of the narrow side of a bench beam, completing a 360 degree turn at mark before dismounting at the end of the bench.
2. *Core Agility*: Complete 4 body shape positions (dish - arch - dish - arch), rotating the body in both directions.
3. *Wobble Spot*: Complete 5 bean bag 'passes' around the body while balancing on the wobble spot on one leg.
4. *Overarm Throw*: Throw a tennis ball, using an overarm throw, at a target approximately 9 metres away.
5. *Basketball Dribble*: Using either hand, dribble a basketball around 4 coloured spots positioned in a 'z' formation.
6. *Catch*: Catch a tennis ball thrown underarm at a rebound net from any distance.
7. *Jumping Patterns*: Complete a jumping pattern sequence that includes a series of hops and jumps (2 footed jump over hurdle > 2 footed landing > 2 left hops > 2 right hops > 2 foot jump over hurdle > 2 footed landing).
8. *T-Agility*: Complete t-agility run, facing forwards throughout.
9. *Sprint*: 10m acceleration to a sprint over finish line.

Data analysis

All data were collected at an acquisition frequency of 40 Hz. The dynamic time warping is implemented on the complete recorded trace, without sectioning or pre-filtering, using the MATLAB function – *'dtw'*. Multidimensional scaling is done using the '*cmdscale*' MATLAB function. Spearman's rank correlation coefficient, r_s, was used to quantify the comparison of human-based and sensor scoring methods. This was calculated using the in-built Matlab function and implemented on the whole activity score. Bland-Altman plots were also used to analyse the level of agreement between human and sensor scores within each individual school-based activity session. Cohen's kappa coefficient, κ, was used to assess machine-human reliability in assignment of performance category (Bronze, Silver, Gold, Platinum). This was calculated with custom algorithms written in Matlab.

3. EXPERIMENTAL PROCEDURES AND RESULTS

Quantification of activity

The complete set of recorded motion data includes 12 time traces, consisting of 3-axis signals from wrist and ankle worn, magnetometer and accelerometer. Even in this small study with 55 participants and using a relatively low sampling frequency of 40 Hz there are over 4 million data points. In this first implementation of the analysis our aim is to demonstrate its applicability using a limited measurement set and so the wrist-worn magnetometer trace for the radial direction (axis running from elbow to wrist) was chosen as the single reporting signal. The aim was to quantify the pattern of movement of each child, over the full set of exercises with a minimised data set, thus allowing for maximal implementation in large cohort studies. Improved accuracy and resolution can of course be obtained by expanding the data set to include orthogonal axes, higher sampling rates, different measurement variables and multiple sensor positions. The positional data from the magnetometer is preferred to the acceleration data as it provides a direct measure of motion through space. The wrist position is chosen as it carries information for all activity types, whereas the ankle mounted sensor shows a limited response for exercises such as throwing or catching. Finally, the radial axis was chosen as it provides a signal that is independent of the compass direction when the arm is in a relaxed position by the side of the body; produces a strong signal for movement within the vertical or horizontal planes and exhibits good differentiation of activities within the exercise sequence.

A typical trace from a child, completing the Dragon challenge within a 2-minute time span, is shown in figure 1. The presence of a signal minimum at points between the activities together with the positive values produced by arm motion during the exercise produces a trace with distinct blocks, corresponding to the 9 discrete activities. Unique signatures corresponding to distinct spatial patterns can be seen, e.g. a series of sharp peaks as a bean bag is passed behind the back during the *wobble spot* activity or a concerted, high displacement band during the bouncing of a ball in the *basketball dribble*.

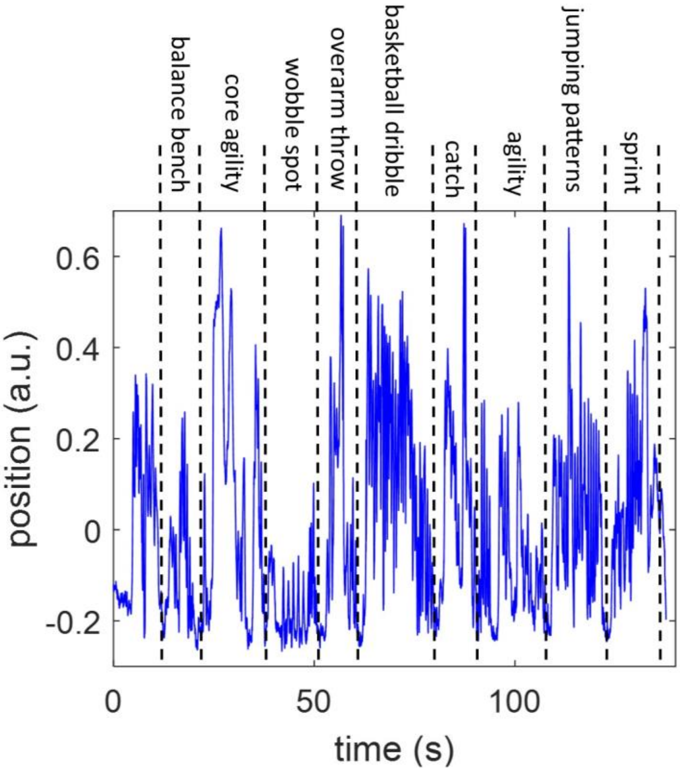

Figure 1: typical time series trace from a wrist-worn sensor with annotated activity sequence. The relative change in position as measured by the magnetometer is shown (arbitrary units).

The appearance of patterns within the magnetometer time trace, that clearly relate to the detail of the required motion within each activity, confirms that the recorded time series is a valid description of a child's performance. The next step in analysis was to compare this time record for different children. This was done by implementing a dynamic time warping on the wrist-magnetometer signals from pairs of participants. A typical pair of traces, prior to and post-implementation of the time warping, is shown in figures 2A & 2B. These are from a platinum category child (above 95th percentile) who completed the challenge in 2 min 14 s (5370 sample pts.) and a slower silver category child (33rd – 66th percentile), who took 2 min 50 s (6780 sample pts.). Implementation of the time warping produces two traces, transformed to be of equal length by the insertion of constant value sequences. This process is completed in a manner that minimises the difference between the two traces. The degree of time warping required is determined by 2 factors: i. the overall time difference for completion of the challenge by the two children (macro-scale) and ii. the point-to-point differences in the signal shape (micro-scale).

The length of the warped time-series is taken as the parameter describing the degree of similarity between a pair of children, i.e. larger differences between traces lead to an enhanced degree of time warping and an extended trace length. This parameterisation can be visually described using a DTW contour map, of the type shown in figures 3A and 3B. The contour map is constructed from a difference matrix $D(t, \tau)$ which describes the normalised difference between signals $S_1(t)$ and $S_2(\tau)$ at

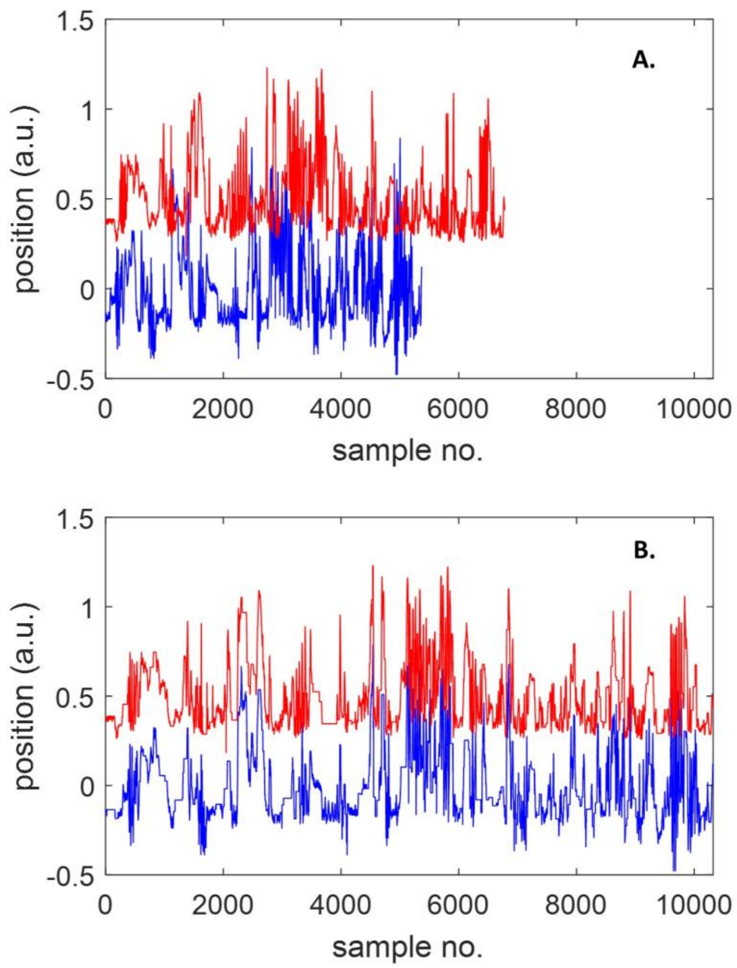

Figure 2: A.) representative, as-measured magnetometer time-series for two children (platinum – blue trace, silver – red trace); B.) the two data sets after implementation of Dynamic Time Warping. (the upper data set in both sub-plots is offset on the y-axis by 0.5 units to provide visual clarity)

time points t and τ:

$$D(t,\tau) = \frac{|S_1(t) - S_2(\tau)|}{max|S_1(t) - S_2(\tau)|_{all\,t,\tau}} \qquad (1)$$

The imaginary line tracing a straight diagonal running from top left to bottom right of the contour map relates the two signals at a common time point and so depicts the signal difference under the condition of $t = \tau$. Thus, for identical signals (or auto-correlation) this will equal zero for all time values.

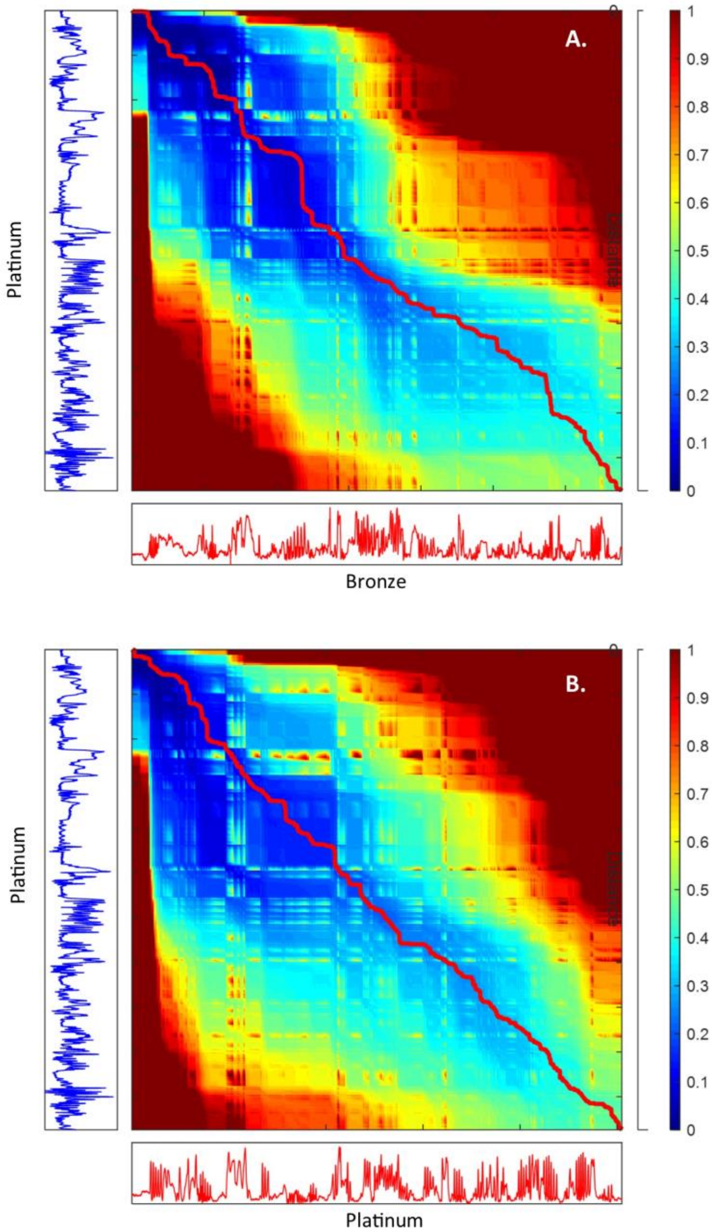

Figure 3: A.) DTW contour map of traces from a Platinum and a Bronze band child; B.) DTW contour map of traces from two Platinum band children. Time-series traces are shown in adjacent sub-plots. The contour scale indicates the signal difference on a normalised scale.

The red lines in figures 3A and 3B trace the actual path of minimum difference; because of non-identical signals this does not follow the diagonal, i.e. for any time point, t a time shift, $\Delta t = |t-\tau|$, is required to find like-values for signals S_1 and S_2. In figure 3A the comparison is between two children with widely differing scores in the challenge. The path of minimum difference therefore deviates markedly from the diagonal, showing large horizontal and vertical steps where the time warping process inserts signal transformations. In contrast, the comparison between like performers in figure

3B produces a much straighter path with only minor deviations from the diagonal. To obtain a quantitative measure of these signal differences, the fractional increase in the minimum difference path length, $\delta(t,\tau)$ relative to the length of the diagonal, $d(t=\tau)$ is taken as the similarity co-efficient, $\Omega_{n,m}$ between children *i* and *j*:

$$\Omega_{i,j} = \frac{(\delta - d)}{d} \quad (2)$$

For a cohort of *n* children, the pairwise comparison by DTW produces an *n* x *n* similarity matrix of $\Omega_{i,j}$ values.

Cohort profiling

To obtain a ranking of all children within a cohort the similarity matrix must transformed to a single parameter data set. This is achieved by clustering the children using multi-dimensional scaling (MDS). The approach allows visualisation of the performance of the cohort and produces a ranking measure that is based on comparison of relative ability between all participants. MDS of an *n* x *n* similarity matrix produces an *n-1* co-ordinate description of the cohort. A 3-D visualisation of cohort performance of 20 children from the same school (school A), based on the first 3 co-ordinates from an MDS analysis, is shown in figure 4. The performance-band of each child is indicated by the colour shading and this indicates that the distribution of the cluster is in the form of a sphere, with the best performers at the centre and progressively lower performers at greater radial distances from the central point.

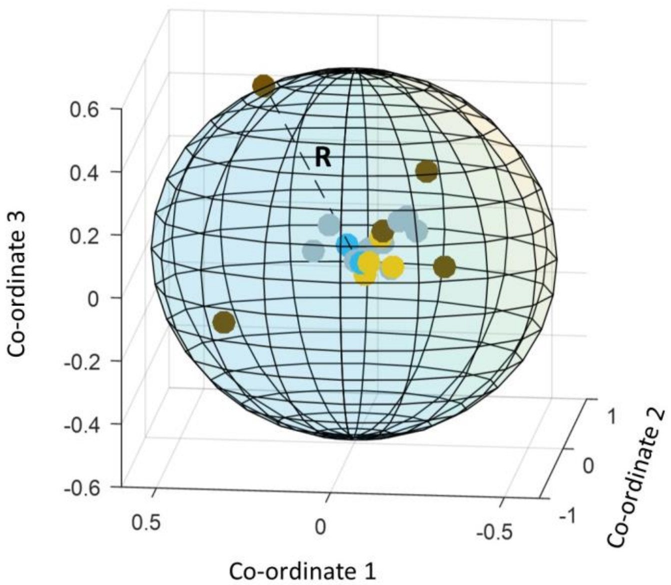

Figure 4: 3D visualisation of a single school cohort based on MDS co-ordinates. Platinum, gold, silver and bronze banding of each child is indicated by colour. Superimposed sphere and scoring radius indicates relative position within the 3-D space. (note: a fixed linear offset is applied to all data points so that the mean value of the platinum-band children is at the [0,0,0] point)

The two platinum-band children from school A are used as a reference standard and included in all data assessments (i.e. they are added to the cohort from other schools). In this way, every similarity matrix incorporates a comparison of participants with these two reference children and so there is benchmarking across all measurements.

Relation of standard and objective assessment

To obtain a score for each child their Euclidean distance, R, is calculated, using all *n-1* co-ordinates of the MDS (when visualising the data within the reduced, 3-dimensional representation, R is the radial distance from the sphere centre). A final transformation is applied to these R-vector values to obtain a sensor score, M which then allows comparison to the human scorer assessments:

$$M = (1.3 - R) \times 50 \tag{3}$$

The (1.3 – R) term inverts the ranking from the MDS to give a maximum rather than a minimum score to the platinum-band children (minimum sphere radius equates to high performance score). The multiplication factor and reference constant within the brackets in equation 3 act to apply a uniform scaling of the dimensionless, MDS vector to match the scoring range of the standardised Dragon Challenge marking scheme. This is necessary as the sensor score stems from a similarity co-efficient whereas the standard scores are based on the Dragon Challenge assessment protocol, in this child performance is scored in situ by trained assessors and recorded on the Child Performance Record form. Their overall score is determined by the total time to completion and skill performance criteria within each exercise (technique and outcome), with each of these given equal weighting. The skill scoring format has a binary structure with a 1 being recorded if a child performs a criterion correctly (e.g. catches a ball) or a 0 if not. The total skills-related score (technique and outcome) for all 9 activities ranges between 0-36 and the time to complete the challenge is scored between 0-18, thus giving a total scoring range of 0-54. To provide a broader comparative measure specific cut-points are generated for the total score using the 33rd, 66th, and 95th percentiles based on pilot data collected across Wales by expert assessors in Spring/Summer 2015. These percentile thresholds were selected to categorise the children into Bronze (lower tertile), Silver (middle tertile), Gold (upper tertile) and Platinum (top 5%) bands. (Full details of the Dragon Challenge scoring protocol can be found in Stratton, 2015).

Correlation plots of sensor-derived and human scores are shown in figures 5A and 5B for a single school (one measurement set) and a group of schools (measurement on multiple days). In both cases, there is a strong, linear correlation between the data sets ($r_s \geq 0.57$, p-value < 0.05)(Cohen 1988), verifying that automated scoring from the sensor data does reflect the judgements made by the human assessors.

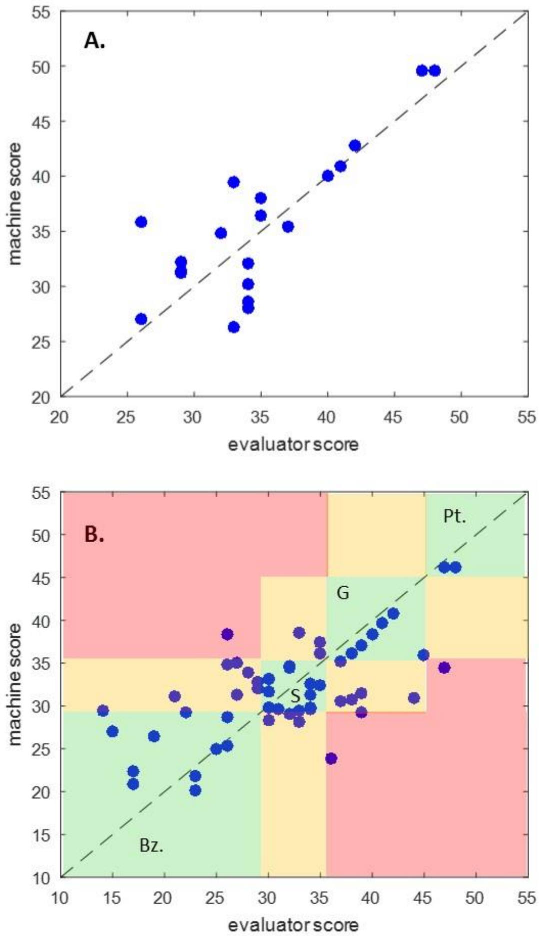

Figure 5: A.) correlation plot of human evaluator and motion sensor derived scores for school A (Spearman correlation co-efficient, $r_s = 0.69$); B.) correlation plot of human evaluator and motion sensor derived scores for children from all 4 participating schools ($r_s = 0.57$). The shaded areas indicate performance-band assignments; green – matched, orange - +/- 1 band, red - +/- 2 bands.

The correlation of performance-band assignments is detailed in Table 1, which presents the classification matrix for band-assignment by sensor and human scores. The banding of the children is also indicated in figure 5B with shaded areas of differing colour indicating regions with a band difference, Δ = 0,1 or 2. The sensor-based categorisation of performance band for children from school A, agreed with the assessors in 65% of cases and within a single band margin of error in 100% of cases (κ-statistic = 0.41). These percentages dropped to 57% and 95% respectively when considering all 55 children, drawn from the 4 schools (κ-statistic = 0.34).

		Sensor-score assignment				
		Bronze	**Silver**	**Gold**	**Platinum**	
Human-score assignment	**Bronze**	15%	5%	2%	0%	22%
	Silver	14%	29%	11%	2%	56%
	Gold	2%	5%	9%	2%	18%
	Platinum	0%	0%	0%	4%	4%
		31%	39%	22%	8%	

Table 1: Classification matrix for performance-band assignment from machine and assessor scores for children from all participating schools.

Bland-Altman plots of the score results from individual schools are shown in figure 6. The correlation co-efficient for a single school cohort ranges between 0.62-0.71, indicating a high level of repeatability for the technique when applied to a single data set in which all children are assessed within a single session. When the data from the 4 schools (4 separate assessment sessions) is combined the correlation co-efficient drops to a value of 0.57. This is not surprising as the data is now subject to the increased variability of multiple studies, each performed in a different setting, on a different day.

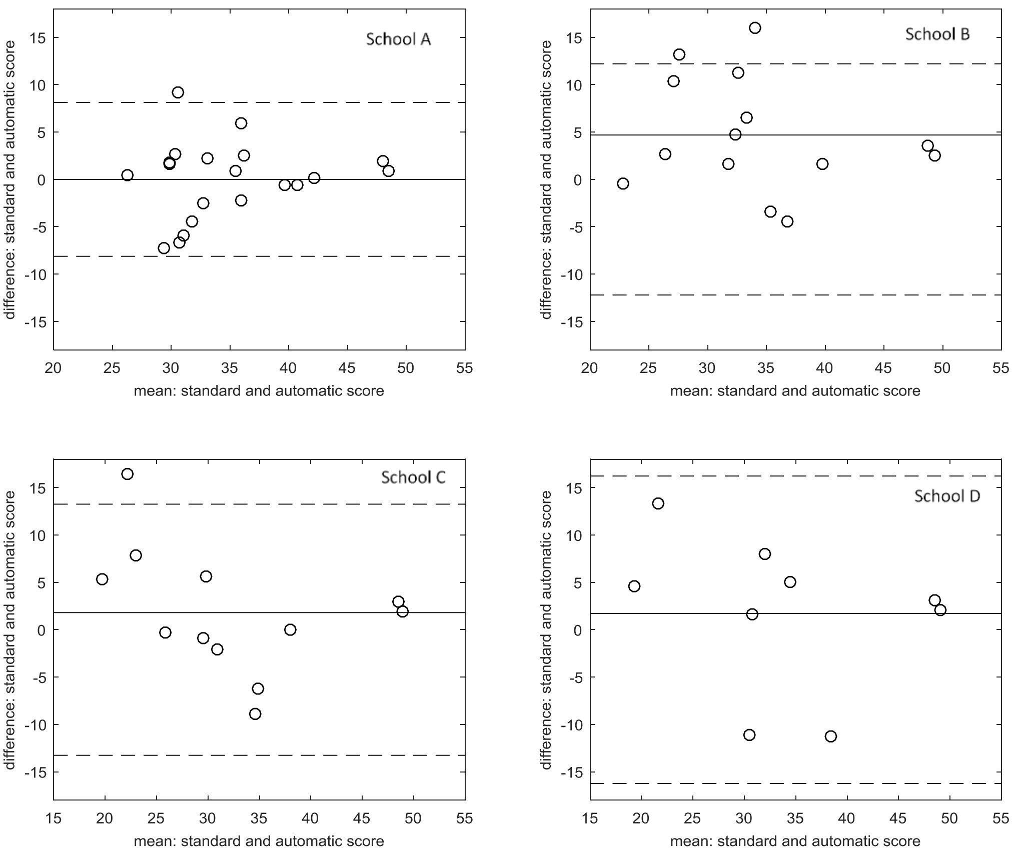

Figure 6: Bland-Altman plots of overall Dragon Challenge score, obtained from standard and automated assessments, for individual schools. Solid line – mean difference; dashed lines - 95% confidence bounds.

4. DISCUSSION AND CONCLUSIONS

In this paper we present an objective, automated approach to assessing fundamental movement skills, using time series analysis on motion sensor data. The technique differs from standard approaches in its delivery of a ranking by comparison rather than by score. The relative indexing of each child's performance is based on a similarity matrix and so depends on all, *n* participants within a study (practically, it depends on the [*n*-1] terms that are summed to give a Euclidean distance). The accuracy of the approach is therefore continually updated as more participants are added, giving an evolving and ever tighter identification of relative performance. The variance in the sensor-based scores, evident in the Bland-Altman plots for each school, relates to the number of assessed children through the similarity matrix. The dimensional scaling gives the most accurate representation possible of all relationships in the matrix. Thus, for the 20-child sample from school A the algorithm is trying to match [n(n-1)]/2 relations (190) whilst for school D with 9 children this becomes 36 relations. As the sample set is increased the performance of a given child is benchmarked against a greater number of their peers and so we obtain greater accuracy is the estimation of their relative ranking in the cohort. This can be seen in the reduction of variance in the score difference in figure 6 plots, for school A.

We choose dynamic time warping in preference to other comparative techniques such as cross-correlation because it highlights the 'time-flow' of the activity. Differences in the pattern or 'flow' of motion over time are quantified by the degree of time warp required to reconcile the data traces. This provides a dual assessment of performance, measuring both macroscale differences in the two, time traces (how fast was the activity completed?) and microscale details of the different motion paths taken (how well was the activity completed?). It is therefore ideally suited to process-oriented assessment of motor competence as it measures the speed with which an activity is completed and parameterises the complex motion patterns that evolve through the exercise. The analysis approach adopted is holistic in nature, being based on correlations of the complete time trace and thus assesses the exercise activity as a whole. This in in contrast to alternative, reductionist cluster-techniques such as principle components analysis, where specific parameterisation from time series features is required. The adoption of a total measure of performance is limiting as we are unable to report on individual skill components such as object control, locomotion or stability(Bisi *et al* 2017). However, the simplification to a single score allows fully automated assessment through the use of a single, wrist mounted sensor.

Multi-dimensional scaling of the similarity matrix provides a visualisation of the performance of the cohort. When imaged in a 3-D space the numerical descriptors of group performance are transformed to a 'performance sphere' - a geometrical construct that positions individuals per ability and contextualises the performance of the whole cohort. Children of like-ability occupy a space defined by the surface of a sphere of given radius. As their performance worsens their distance from the centre of the sphere increases, i.e. they are further away from the optimum. Poorer ability children also exhibit greater variability in performance; in the visual representation, this relates to the increased surface area produced by a greater sphere radius. At a given radius, differences in the position of children on the sphere surface tells us about the different strategies adopted in performing the tasks – these children have similar overall performance (same radius on sphere) but can be separated by large distances if their motion patterns show marked dissimilarity. The geometry of the spherical performance map indicates a narrowing of the FMS 'window' for better performers: there are multiple ways in which performance can be poorer (e.g. within different activities) whereas high performance children tend towards an optimal technique.

Comparison of the automated, sensor-based method to the standard approach indicates a strong correlation between absolute scores (r_s range: 0.57 to 0.71) and a moderate correlation for categorical classification of performance (κ range: 0.34 to 0.41). Comparative studies on measurement variability within physical activity tests have reported r_s values of 0.6 when comparing overall scores between different FMS tests(Lander *et al* 2017) and $r_s \sim 0.5 – 0.7$ for comparison of process and product-oriented scores of individual skills(Logan *et al* 2017). Moreover, comparison of inter-rater variability within a single test indicated κ values in the range of 0.2 – 0.6 for overhand throw and strike skills(Barnett *et al* 2014). Thus our approach would appear to provide a robust measurement tool that is as accurate as can be expected given these limits in the ability of any single exercise programme to fully capture the intricacy of motor skills and the innate variability of human scorers.

Comparison of sensor and human assessment is required to benchmark the automated approach and so assess applicability. However, this is a comparison of two markedly different methodologies and so it is not appropriate to ask the question of 'which is right and which is wrong?' Rather, a common goal is approached from two different directions: i. absolute scoring of individuals and individual movement components and ii. relative ranking of individuals from similarity measurement of the whole activity. The question we seek to answer is 'do these different techniques give a common result when profiling a group's activity?', our results indicate that indeed they do. Having established a commonality of result the sensor-based assessment can be used as a powerful and robust adjunct to assessment by expert scorers. Sensor-based scoring provides a check and common reference, allowing cross-validation of scoring tables between assessors. From this, adjusted weighting of scores between schools and across wider geographical areas can be implemented, to correct for in-built bias and systematic error. For example, within this study the Bland-Altman plots in figure 6 show a significant bias in the data from school B. Given that all the score sets are consistently moderated through the inclusion of the two Platinum-band performers the appearance of a marked bias in this data set raises questions as to the validity of this activity session. The introduction of objective scoring with in-built checks such as this, opens the possibility of using untrained assessors (e.g. school teachers, sports coaches), allowing wider deployment of the activity programme as accurate and dependable ranking is provided by sensor-measurement.

The development of automated assessment techniques for fundamental movement skills will augment outcome measures used in intervention programmes, designed to promote physical activity and improve physical literacy. Longitudinal assessments are therefore required so that the on-going improvement in children's physical ability can be captured. The quantitative, sensor-based profiling presented here is ideally suited to this task. It is robust, accurate and objective, but also, most importantly it is based on comparison and so the evolving performance of an individual is easily referenced to previous assessments. This is best envisaged in the context of the 'performance sphere', where the activity of a child over a series of assessments would be visualised as an evolving path through the sphere; mapping their personal performance history and moving ever closer to the centre as movement skill levels improved relative to their previous capability and that of their peers.

ACKNOWLEDGEMENTS

Claire Barnes was supported during this work by a PhD scholarship from the Engineering and Physical Sciences Research Council, U.K.

5. REFERENCES

Akl A, Feng C and Valaee S 2011 A novel accelerometer-based gesture recognition system *IEEE Trans. Signal Process.* **59** 6197–205

Allen F R, Ambikairajah E, Lovell N H and Celler B G 2006 Classification of a known sequence of motions and postures from accelerometry data using adapted Gaussian mixture models *Physiol. Meas.* **27** 935–51

Andrew N E, Gabbe B J, Wolfe R and Cameron P A 2010 Evaluation of instruments for measuring the burden of sport and active recreation injury *Sport. Med.* **40** 141–61

Barnes C M, Clark C C T, Holton M D, Stratton G and Summers H D 2016 Quantitative Time-Profiling of Children's Activity and Motion *Med. Sci. Sports Exerc.*

Barnett L M, Minto C, Lander N and Hardy L L 2014 Interrater reliability assessment using the Test of Gross Motor Development-2 *J. Sci. Med. Sport* **17** 667–70

Bisi M C, Pacini Panebianco G, Polman R and Stagni R 2017 Objective assessment of movement competence in children using wearable sensors: An instrumented version of the TGMD-2 locomotor subtest *Gait Posture* **56** 42–8

Borg I and Groenen P 2005 *Modern Multidimensional Scaling - Theory and Applications* (New York: Springer)

Brandes M, Zijlstra W, Heikens S, van Lummel R and Rosenbaum D 2006 Accelerometry based assessment of gait parameters in children *Gait Posture* **24** 482–6

Braun E, Geurten B and Egelhaaf M 2010 Identifying prototypical components in behaviour using clustering algorithms *PLoS One* **5**

Coe D P, Pivarnik J M, Womack C J, Reeves M J and Malina R M 2006 Effect of physical education and activity levels on academic achievement in children *Med. Sci. Sports Exerc.* **38** 1515–9

Cohen J 1988 Statistical power analysis for the behavioral sciences *Stat. Power Anal. Behav. Sci.* **2nd** 567

Ermes M, Pärkkä J, Mäntyjärvi J and Korhonen I 2008 Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions *IEEE Trans. Inf. Technol. Biomed.* **12** 20–6

Fisher A, Reilly J J, Kelly L A, Montgomery C, Williamson A, Paton J Y and Grant S 2005 Fundamental movement skills and habitual physical activity in young children *Med. Sci. Sports Exerc.* **37** 684–8

Godfrey A, Conway R, Meagher D and ÓLaighin G 2008 Direct measurement of human movement by accelerometry *Med. Eng. Phys.* **30** 1364–86

Halliday S E, Winter D A, Frank J S, Patla A E and Prince F 1998 The initiation of gait in young, elderly, and Parkinson's disease subjects *Gait Posture* **8** 8–14

Hausdorff J M, Cudkowicz M E, Firtion R, Wei J Y and Goldberger A L 1998 Gait variability and basal ganglia disorders: Stride-to-stride variations of gait cycle timing in Parkinson's disease and Huntington's disease *Mov. Disord.* **13** 428–37

Hills A P, Andersen L B and Byrne N M 2011 Physical activity and obesity in children. *Br. J. Sports Med.* **45** 866–70

Ten Holt G a, Reinders M J T and Hendriks E a 2007 Multi-Dimensional Dynamic Time Warping for Gesture Recognition *Time* **5249** 23–32

Hout M C, Papesh M H and Goldinger S D 2013 Multidimensional scaling *Wiley Interdiscip. Rev. Cogn. Sci.* **4** 93–103

Janz K F 2006 Physical activity in epidemiology: moving from questionnaire to objective measurement *Br. J. Sports Med.* **40** 191–2

Lander N, Morgan P J, Salmon J, Logan S W and Barnett L M 2017 The reliability and validity of an authentic motor skill assessment tool for early adolescent girls in an Australian school setting *J. Sci. Med. Sport* **20** 590–4

Logan S W, Barnett L M, Goodway J D and Stodden D F 2017 Comparison of performance on process- and product-oriented assessments of fundamental motor skills across childhood *J. Sports Sci.* **35** 634–41

Lohman T G, Roche A F and Martorell R 1988 Anthropometric standardization reference manual *Hum. Kinet. Books* 177

Lynch E, Hillier S and Cadilhac D 2014 When should physical rehabilitation commence after stroke: a systematic review. *Int. J. Stroke* **9** 468–78

Mannini A, Intille S S, Rosenberger M, Sabatini A M and Haskell W 2013 Activity recognition using a single accelerometer placed at the wrist or ankle *Med. Sci. Sports Exerc.* **45** 2193–203

Mathie M J, Coster A C F, Lovell N H and Celler B G 2004 Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement. *Physiol. Meas.* **25** R1-20

Maud P J and Foster C 2006 *Physiological assessment of human fitness* (Human Kinetics)

Morgan P J, Barnett L M, Cliff D P, Okely A D, Scott H A, Cohen K E and Lubans D R 2013 Fundamental Movement Skill Interventions in Youth: A Systematic Review and Meta-analysis *Pediatrics* **132** e1361–83

Musculo R, Conforto S, Schmid M, Caselli P and D'Alessio T 2007 Classification of motor activities through derivative dynamic time warping applied on accelerometer data *Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc.* 4930–3

Myer G D, Faigenbaum A D, Edwards N M, Clark J F, Best T M and Sallis R E 2015 Sixty minutes of what? A developing brain perspective for activating children with an integrative exercise approach. *Br. J. Sports Med.* 1–9

Penedo F J and Dahn J R 2005 Exercise and well-being: a review of mental and physical health benefits associated with physical activity *Curr. Opin. Psychiatry* **18** 189–93

Podlog L and Eklund R C 2006 A longitudinal investigation of competitive athletes' return to sport following serious injury *J. Appl. Sport Psychol.* **18** 44–68

Robinson L E, Stodden D F, Barnett L M, Lopes V P, Logan S W, Rodrigues L P and D'Hondt E 2015 Motor Competence and its Effect on Positive Developmental Trajectories of Health *Sport. Med.* **45** 1273–84

Rudd J, Butson M L, Barnett L, Farrow D, Berry J, Borkoles E and Polman R 2016 A holistic measurement model of movement competency in children *J. Sports Sci.* **34** 477–85

Sallis J F, Prochaska J J and Taylor W C 2000 A review of correlates of physical activity of children and adolescents. *Med. Sci. Sport. Exerc.* **32** 963–75

Sirard J R and Pate R R 2001 Physical activity assessment in children and adolescents. *Sport. Med.* **31** 439–54

Stratton G, Foweather L, Rotchell J, English J and Hughes H 2015 *Dragon Challenge V1.0 Manual* (Cardiff: Sport Wales)

Tremblay M and Lloyd M 2010 Physical Literacy Measurement - The Missing Piece. *Phys. Heal. Educ. J.* **76** 26–30

Trost S G 2001 Objective measurement of physical activity in youth: current issues, future directions. *Exerc. Sport Sci. Rev.* **29** 32–6

Zhou F and De La Torre F 2016 Generalized Canonical Time Warping *IEEE Trans. Pattern Anal. Mach. Intell.* **38** 279–94