Multisite accelerometry for sleep and wake classification in children

2015 Physiol. Meas. 36 133

(http://iopscience.iop.org/0967-3334/36/1/133)

IOP Publishing | Institute of Physics and Engineering in Medicine Physiological Measurement

Physiol. Meas. **36** (2015) 133–147 doi:10.1088/0967-3334/36/1/133

Multisite accelerometry for sleep and wake classification in children

Marnie L Lamprecht[1], Andrew P Bradley[1], Tommy Tran[2], Alison Boynton[2] and Philip I Terrill[1]

[1] The University of Queensland, School of Information Technology and Electrical Engineering, Brisbane, St Lucia QLD 4072, Australia
[2] Department of Respiratory and Sleep Medicine, The Mater Children's Hospital, Raymond Terrace, South Brisbane QLD 4101, Australia

E-mail: m.lamprecht@uq.edu.au

Received 10 July 2014, revised 10 September 2014
Accepted for publication 13 October 2014
Published 16 December 2014

Abstract

Actigraphy is a useful alternative to the gold standard polysomnogram for non-invasively measuring sleep and wakefulness. However, it is unable to accurately assess sleep fragmentation due to its inability to differentiate restless sleep from wakefulness and quiet wake from sleep. This presents significant limitations in the assessment of sleep-related breathing disorders where sleep fragmentation is a common symptom. We propose that this limitation may be caused by hardware constraints and movement representation techniques. Our objective was to determine if multisite tri-axial accelerometry improves sleep and wake classification. Twenty-four patients aged 6–15 years (median: 8 years, 16 male) underwent a diagnostic polysomnogram while simultaneously recording motion from the left wrist and index fingertip, upper thorax and left ankle and great toe using a custom accelerometry system. Movement was quantified using several features and two feature selection techniques were employed to select optimal features for restricted feature set sizes. A heuristic was also applied to identify movements during restless sleep. The sleep and wake classification performance was then assessed and validated against the manually scored polysomnogram using discriminant analysis. Tri-axial accelerometry measured at the wrist significantly improved the wake detection when compared to uni-axial accelerometry (specificity at 85% sensitivity: 71.3(14.2)% versus 55.2(24.7)%, $p < 0.01$). Multisite accelerometry significantly improved the performance when compared to the single wrist placement (specificity at 85% sensitivity: 82.1(12.5)% versus 71.3(14.2)%, $p < 0.05$). Our results indicate that multisite accelerometry offers a significant performance benefit which could be further improved by analysing movement in raw multisite accelerometry data.

Keywords: accelerometer placement, actigraphy, sleep assessment

(Some figures may appear in colour only in the online journal)

1. Introduction

Actigraphy is a useful tool for non-invasively identifying sleep and wakefulness for the assessment of sleeping disorders. Commercial sleep systems typically use an accelerometer located on the wrist to measure movement. Movement is then quantified as a summarised *activity count* within a fixed time-frame, or epoch (typically 30 s) using several time-series methods. A threshold-based classifier is then used to identify periods of inactivity as sleep and active periods as wake. Despite achieving greater than 90% agreement with the gold standard polysomnogram (Jean-Louis *et al* 2001), commercial systems will often misclassify periods of quiet wake as sleep and periods of restless sleep as wake, resulting in poor specificity (commonly 40–60%) (Sadeh 2011, Meltzer *et al* 2012). While this is acceptable for the assessment of circadian rhythms, it presents significant limitations in disorders such as obstructive sleep apnoea, where accurate quantification of sleep fragmentation is an important diagnostic consideration (Roebuck *et al* 2014). Consequently, there is significant motivation to identify technological advancements to improve the performance of actigraphy for assessing sleep and wakefulness in non-ideal patients.

Formerly, commercial sleep actigraphy systems measured activity using a uni-axial accelerometer located at the wrist. The inability of uni-axial accelerometry to identify movements along other axes has motivated recent commercial sleep systems to employ tri-axial accelerometry. While literature has compared the ability of different commercial systems to estimate sleep parameters, there have been no known comparisons of uni-axial and tri-axial accelerometry using the same device. Furthermore, measuring movement solely on the wrist restricts the identification of movements that occur elsewhere on the body (for example, leg movements or the intensity of body positional changes). This limitation is likely to contribute to the poor specificity of conventional systems as certain movements will go undetected. Consequently, there have been a number of attempts in literature to identify the efficacy of replacing the conventional wrist accelerometer placement with waist, hip or shoulder placements for detecting sleep and wakefulness in children (Maija-Riikka Steenari and Aronen 2002, Adkins *et al* 2012, Kinder *et al* 2012), infants (Sazonov *et al* 2004) and adults (Van Hilten *et al* 1993, Middelkoop *et al* 1997, Rapp *et al* 2010). These studies have found that changing the accelerometer placement offers similar, if not better, performance than the wrist. Continuing this analysis to include multiple accelerometer placements may further improve the performance. Indeed, literature in physical activity assessment and task identification have found that combining information from multiple sensors does improve the performance (Bao and Intille 2004, Thiemjarus *et al* 2004). Despite this, literature attempting to identify the efficacy of combining multiple accelerometer placements for assessing sleep and wakefulness in children relative to the polysomnogram is lacking.

Another potential cause of the poor specificity inherent in conventional sleep actigraphy systems is the inability to detect restless sleep. This is because actigraphy will often misclassify movements associated with sleep as having occurred during wakefulness, resulting in false wake detections. However, there have been few attempts to differentiate movements during restless sleep from those during wakefulness (Domingues *et al* 2012, Domingues *et al* 2013), with most literature focusing on movement duration and frequency (Giganti *et al* 2008, Gori *et al* 2004). While these techniques summarise the occurrence of activity well, they do not provide an acceptable resolution for identifying specific characteristics of movements. A different approach

to identifying restless sleep is that employed by Crespo *et al* (2012). They identified periods of activity and inactivity during sleep using a stochastic approach to assess the likelihood of a conventional activity count corresponding to sleep or wake based on its surrounding epochs. While this approach was effective at detecting restless sleep in patients without fragmented sleep, it is likely to falsely identify short periods of wake as restless sleep. An alternative approach which may overcome these limitations is to differentiate movements associated with restless sleep from wake by characterising the nature of these movements as observed in raw accelerometry data.

The specific aim of this study was therefore to determine whether the additional information contained in multisite tri-axial accelerometry enhances the sleep and wake classification ability of actigraphy relative to conventional accelerometry of the wrist in children. To address this aim we compare the performance of (1) uni-axial and tri-axial accelerometry measured at the wrist; (2) classifying sleep and wakefulness based on combined information from multiple accelerometer placements (left index fingertip, left wrist, upper thorax, left ankle and left great toe) relative to single site accelerometry; and (3) heuristically identifying restless sleep by analysing raw accelerometry data as a preliminary processing step.

2. Method

2.1. Patient characteristics

Pediatric patients attending the Mater Children's Hospital for a full overnight diagnostic polysomnogram (PSG) to investigate suspected sleep disordered breathing were invited to participate in our study. Patients were enrolled as participants if their guardians gave written consent and the patient gave verbal consent. In total, 33 patients aged 5–16 years (median 9 years, 23 male) with obstructive sleep apnoea (OSA) severity ranging from healthy to severe (median AHI 1.15, range 0–16.9) were recruited in this study.

2.2. Data acquisition

In conjunction with full overnight diagnostics polysomnography (PSG) (Beck and Marcus 2009), all patients wore a custom Continuous Multisite Accelerometry System (CMAS) developed specifically to allow analaysis of high resolution, tri-axial raw accelerometry data at multiple locations on the body, as previously described by Terrill *et al* (2010). CMAS records raw tri-axial 8-bit accelerometry data (range of $\pm 2g$, where g represents the acceleration due to gravity) with a 100 Hz sampling rate at the left wrist and left index fingertip, upper thorax, left ankle and left great toe. Data were gathered from these locations as a reasonable trade-off between capturing the majority of movement types (for example, limb and extremity twitches, rolls, positional changes, etc) and minimizing the number of sensor modules on the patient. Data were transmitted wirelessly to a receiver unit and logged on a personal computer, separate to the PSG system. To enable post-study synchronisation with the PSG, the receiver unit also logged ankle and toe data as additional traces in the PSG montage, through an analog input to the PSG system. Figure 1 describes the hardware components of the custom accelerometry system. To provide a preliminary comparison with a commercial system, all patients also wore the Actiwatch Mini (CamNTech) on the left wrist.

2.3. Pre-processing

All data were manually pre-screened for artefacts caused by malfunctioning accelerometers and manually synchronised with the PSG recording using custom software. Nine patients

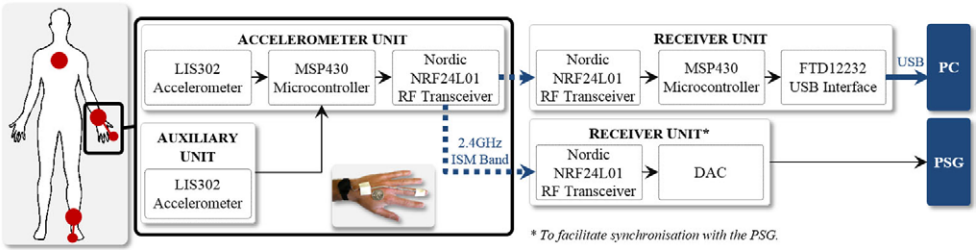

Figure 1. Hardware components of the continuous multisite accelerometry system (CMAS). Shown here is the wrist unit with the finger auxiliary unit. The wrist, upper thorax and ankle modules are the main accelerometry units. The finger and toe modules are auxiliary units that are connected to the wrist and ankle units respectively.

were unable to be used in analysis due to hardware malfunction, incorrect clinical setup or patient interference. A total of 24 studies (aged 6–15 years, median 8 years, 16 male) were satisfactorily completed and able to be synchronised for further analysis. Blocks of missing data of up to 2 s in duration (caused by missing wireless packets) were identified and interpolated using piecewise cubic interpolation. The interpolation was calculated using the surrounding 0.1 s of accelerometry data. Coupled with the 2 s duration restriction, this ensures that no artificial 'movements' were introduced by the interpolation procedure. In order to extract information summarising postural changes (i.e. the DC component of accelerometry) and transient motion (i.e. higher frequency components), two separate filter protocols were applied: (1) a 5th order low pass (<2 Hz) Butterworth filter (LPF); and (2) a 10th order band-pass (2–12 Hz) Butterworth filter (BPF). The 12 Hz cutoff frequency was chosen on the basis of sample spectral analysis of raw accelerometry during typical movements, which indicated that information content predominantly occurred at frequencies lower than 12 Hz. The filtered 100 Hz raw accelerometry data were down-sampled to 40 Hz to improve analysis speed while avoiding aliasing artefacts.

2.4. Feature extraction

To ensure interpretability in the context of existing literature and clinical practice, key clinical conventions were adopted in this study. Firstly, we use 30 s windows of data to correspond with the 30 s sleep scoring class labels (as defined by the PSG). Second, we adopt windowed time-series features as commonly utilised in commercial sleep actigraphy systems (time above threshold, TAT; digital integration, DI; and zero-crossing, ZC) (Ancoli-Israel *et al* 2003). For the tri-axial accelerometry data, these features were derived using the Euclidean norm of the band-pass filtered tri-axial data $\| \mathbf{a}_B \|$,

$$\mathrm{TAT} = \sum_{n=1}^{N \cdot f_s} \mathbb{1} \{ \| \mathbf{a}_B [n] \| > T \} , \tag{1}$$

$$\mathrm{DI} = \sum_{n=1}^{N \cdot f_s} \| \mathbf{a}_B [n] \|, \tag{2}$$

$$\mathrm{ZC} = \sum_{n=1}^{N \cdot f_s} \mathbb{1} \{ (\| \mathbf{a}_B [n] \| - T) \cdot (\| \mathbf{a}_B [n-1] \| - T) < 0 \} , \tag{3}$$

where $N = 30$ s represents the window size, $f_s = 25$ Hz represents the sampling rate, $\mathbb{1}$ represents the indicator function and T represents a threshold. These simple representations of motion form a sensible starting point for this analysis. Further, while there are numerous alternative features, the conventional features allow our results to be generalised in the context of commercial systems for the purpose of addressing the key study question, i.e. the value of multisite tri-axial accelerometry relative to single site accelerometry. In order to fully exploit the nature of tri-axial accelerometry, two additional time-series features were extracted from CMAS: the maximum magnitude of acceleration (MAXACT) and the integrated angle of postural change (SUMPST). MAXACT represents limb twitches and abrupt movements by measuring the maximum acceleration of the limb within each epoch. The band-pass filter was applied to isolate the higher frequency motion $\mathbf{a}_B$ used to derive MAXACT,

$$\text{MAXACT} = \max\left(\forall\, n \in N{\cdot}f_s \,:\left\|\, \mathbf{a}_B[n] \,\right\|\right). \tag{4}$$

SUMPST represents limb postural changes and slow motion by approximating the total angular displacement of the limb relative to gravity between consecutive samples within each epoch. The low-pass filter was applied to the raw accelerometry data to isolate the low frequency motion $\mathbf{a}_L$ used to derive SUMPST,

$$\text{SUMPST} = \sum_{n=1}^{N\cdot f_s} \left|\cos^{-1}\left(\frac{\mathbf{a}_L[n]\cdot\mathbf{a}_L[n+1]}{\|\mathbf{a}_L[n]\|\cdot\|\mathbf{a}_L[n+1]\|}\right)\right|. \tag{5}$$

2.5. Feature selection

We use two well established feature selection techniques in this analysis to judge the reliability and generalisability of the selected features: (1) a sequential forward selection feature search optimising for partial receiver operating characteristic area under curve (PAUC) (Dodd and Pepe 2003) and (2) Minimum Redundancy Maximum Relevancy (mRMR) (Yu and Liu 2004). We also use (3) a heuristic on the raw accelerometry data to detect and effectively negate the influence of movements that occur during restless sleep, prior to generating the features. Each feature selection method employs leave-one-out cross-validation on subjects to select the features.

In the first approach, we use a forward selection technique for our feature search. This technique aims to maximize the area under the ROC curve above 60% specificity. This restriction attempts to maximize the ability of the features to discriminate between sleep and wakefulness. However, it is possible that this search can select redundant features, which may limit the performance when restricting the size of the feature set (Cios *et al* 1998). The second method addresses this by minimising the redundancy in the selected features while maximising the relevancy to the class label. The third technique aims to reduce the number of misclassifications caused by restless sleep. The three methods are described below.

2.5.1. Forward selection with partial AUC. The starting point for our analysis is a simple, commonly used feature selection technique. Sequential forward feature selection performs a greedy search for features that improves a performance metric (Liu and Motoda 1998). The search finishes when the metric is no longer improved by additional features. The performance metric we use in this study is the partial area under the receiver operating characteristic curve (PAUC) (Dodd and Pepe 2003). Here PAUC is defined as the AUC above a specificity of 60%. AUC represents the ability of a feature or feature set to rank a randomly chosen positive

instance higher than a randomly chosen negative instance (Bradley 1997). Therefore, maximizing AUC above a specificity of 60% will effectively maximize the sleep and wake discriminatory power of the specific features. Limiting the area in this way ensures that a feature set that performs significantly well at a high specificity will be chosen. While this addresses the issue of low specificity (Meltzer *et al* 2012), it is possible that redundant features can be selected. Reducing redundancy in the selected features can improve generalisability (Cios *et al* 1998, Mitra *et al* 2002). This is addressed by the next method which seeks to minimise redundancy.

2.5.2. Minimum redundancy maximum relevancy. Minimum redundancy maximum relevancy (mRMR) is a supervised feature selection technique that aims to minimize the redundancy between features while maximizing the relevancy to the class label (Yu and Liu 2004). This method addresses a fundamental problem with the maximum relevancy feature selection technique (Saeys *et al* 2007); features can be highly relevant to the class label but also dependent on each other, thus giving a larger feature set than required. mRMR is particularly suited to analysing movements from different locations, as this data is likely to be somewhat redundant. We use a similar method to Peng *et al* (2005) to define and optimize the redundancy and relevancy of the feature. The redundancy *Ru* of a feature set is defined as the mean mutual information *I* between each feature f_n within that set *s*,

$$Ru = \frac{1}{|s|^2} \sum_{f_1, f_2 \in s} I\left(f_1; f_2\right), \tag{6}$$

where mutual information is defined by Hutter (2002) as,

$$I(s; y) = \sum_{y \in Y} \sum_{f \in s} \log \left(\frac{p(f, y)}{p(f) \cdot p(y)} \right), \tag{7}$$

where $p(f, y)$ is the joint probability distribution function of the class *y* and the feature *f* and $p(f)$ and $p(y)$ are the marginal probability distribution functions of the feature and the class respectively. Mutual information quantifies the contribution each feature has to the class prediction and the similarity of each feature to each other. Similar to the redundancy, the relevancy *Re* is defined as the mean mutual information *I* between each feature *f* within a feature set *s* and the class label *c*,

$$Re = \frac{1}{|s|} \sum_{f \in s} I(f; c). \tag{8}$$

A simple approach to optimize for maximum relevancy and minimum redundancy is to maximize the difference between the two metrics,

$$\text{mRMR} = \max\ (Re - Ru). \tag{9}$$

2.5.3. Heuristic removal of restless sleep. A key limitation of conventional actigraphy is the inability to detect restless sleep and quiet wake. This inability to differentiate movements associated with sleep from those associated with wakefulness results in false positive detections and consequently poor specificity. This limitation may be due to the conventional movement representation whereby activity is summarised over a fixed (typically 30 s) time frame. Because of this low resolution, activity counts are unable to determine what type of activity occurred and as such may be unable to exploit the different characteristics of

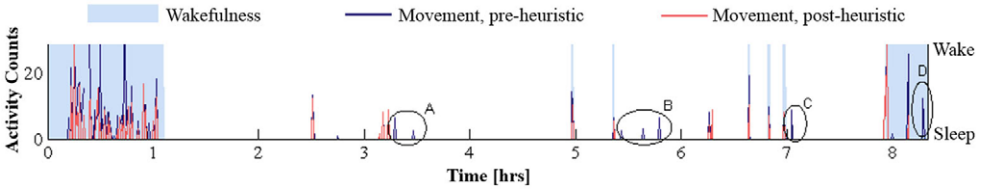

Figure 2. Example of the effect of applying the heuristic to remove restless sleep for one patient. Sleep and wake (shaded) periods are shown, where sleep is defined as all non-REM and REM stages. Movement is represented by the zero-crossing method for the wrist. The heuristic removed movement corresponding to restless sleep (A–C). Movements during restless sleep that were not removed were likely movements that also occurred during wakefulness and were consequently ignored by the heuristic. The heuristic also removed some movements corresponding to wake (D) as there is a small likelihood that movements corresponding to restless sleep can also occur during wake.

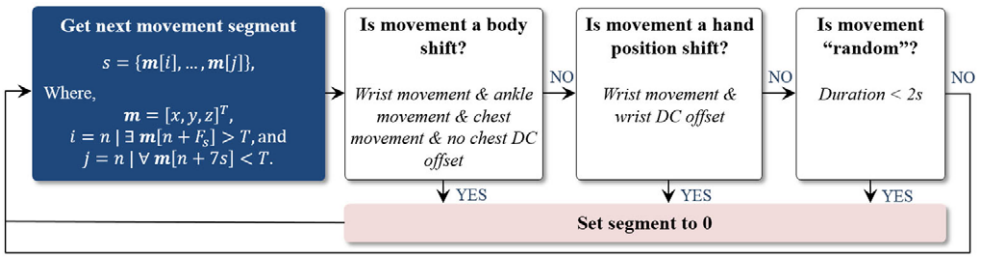

Figure 3. The process for removing restless sleep from the raw accelerometry data; data is zeroed if the segment is found to be a body shift, hand position change or a 'random' movement. A movement segment is defined as any datapoints that are above the noise floor until there is only the noise floor for at least 7 s.

movement. Prior literature has focused on detecting periods of activity and inactivity. However, we analysed segments of videos of sleeping patients and found that there may be some benefit to detecting different *types* of activity. That is, we noted that there were characteristics of some movements that were more likely to occur during sleep than during wakefulness. In particular, we observed that children were more likely to (a) completely change their hand positions during sleep and (b) slightly shift their body. We also found that (c) 'random' movements during sleep were likely to be very short in duration (less than 2 s). Conversely, movements such as face scratching and full body position changes were likely to occur during both sleep and wake and movements associated with wake were likely to have a longer duration. Therefore, we postulate that detecting different types of movement may improve sleep and wake classification performance (an example of the effect of identifying these movements is illustrated in figure 2).

To initially investigate this, we adopt a heuristic whereby we identify and quantify the observed characteristics of movements associated with sleep (described as (a)–(c) above) and effectively negate their influence on the raw accelerometry data before generating the features. That is, we set the raw data corresponding to these segments to zero so that the classifier will see 'no movement' and consequently label these segments as *sleep*. The process is described in figure 3. Movements are detected from the raw accelerometry data as any datapoints that are above the noise floor. As the magnitude of the noise is consistent across the raw signal, the noise floor is defined as the median of the absolute magnitude of the high-pass filtered

accelerometry data across each full study. The duration is defined as the start of movement until there is no movement for a minimum of 7 s. This constraint ensures that transitions of movements are detected as a single 'movement', which is particularly important for movements such as sleeping position changes, e.g. supine to left or right lateral. We define the characteristics of movements that we observed during sleep as (a) hand position change: wrist movement with a final wrist DC offset; (b) body shift: wrist, ankle and chest movement with no chest DC offset; and (c) 'random' movement: any movement of less than two seconds in duration.

2.6. Feature validation procedure

The sleep and wake predictive performance of the selected features was determined using quadratic discriminant analysis (QDA) with a binomial distribution in a leave-one-out cross-validation design on subjects. We used QDA as the variability of the accelerometry data for sleep and wake differ, which violates the assumptions for a linear discriminant (Lachenbruch and Goldstein 1979). The probability of the predicted sleep stages given by the discriminant analysis was compared to the actual sleep stages using receiver operating characteristic (ROC) curve analysis. The ability of the predictor to discriminate between sleep and wake was summarised by the ROC area under curve (AUC) (Bradley 1997). The predictive performance at each fold of the cross-validation was quantified using Kappa agreement (κ) with the PSG (Feinstein and Cicchetti 1990), where wake was defined as a predictive probability greater than the threshold that gave the maximum Kappa agreement in the ROC analysis. The primary outcome measure in our analysis is the specificity at a fixed sensitivity. This metric summarises the ability of the features to accurately detect wakefulness without compromising sleep detection.

We note that agreement with the polysomnogram is not an effective representation of the predictive accuracy. As the prior probabilities of sleep and wakefulness across the night are skewed towards sleep, agreement with the polysomnogram is positively biased and therefore not reflective of the true performance (Domingues *et al* 2014). For this reason, we have included an unbiased estimate of accuracy (κ) and the class discrimination summary metrics (AUC and specificity).

2.7. Statistical analysis

Most metrics exhibited equal variance (as defined by the Brown–Forsythe test). Therefore two-way ANOVA was used to assess the effect of the subjects and the prediction from QDA for the selected features. Post hoc Tukeys Honestly Significant Differences (HSD) test was then used to compare the means of any significant main effect defined by the ANOVA. Welchs t-test was used where unequal variance was observed.

3. Results

Of the 33 patients recruited in this study, 24 patients aged 6–15 years (median 8 years, 16 male) with OSA severity ranging from healthy to severe (median AHI 1.15, range 0–16.9) were eligible for the full analysis in this study; nine subjects aged 8–16 years (median 12.5 years, 7 male) with OSA severity ranging from healthy to moderate (median AHI 0.4, range 0–7.7) were excluded due to technical issues with actigraphy. The ROC analysis and agreement with the gold standard PSG are summarised for the best N features of each of

Table 1. Class discriminibility and predictive performance for the N best features of each feature selection method.

N	Selected features	Sp [%] (at 85% Se)	AUC [%]	Agreement [%]	κ
	Uni-axial, $Wrist_{ZC}$	55.2 (24.7)	79.9 (8.6)	84.0 (6.4)	0.464 (0.155)
1	Tri-axial, $Wrist_{ZC}$	71.3 (14.2)[b]	83.3 (7.5)	83.1 (7.7)	0.495 (0.183)
Multivariate Locations, Forward Selection with PAUC					
2	$Wrist_{ZC}$, Toe_{ZC}	72.7 (21.6)[a]	85.8 (8.6)[b]	85.9 (8.4)	0.549 (0.206)[a]
3	$Wrist_{ZC}$, Toe_{ZC}, Toe_{TAT}	73.0 (21.1)[a]	86.1 (8.6)[b]	86.2 (8.4)	0.553 (0.209)[a]
⩽5	$Wrist_{ZC}$, Toe_{ZC}, Toe_{TAT}, $Ankle_{DI}$, Toe_{MAXACT}	79.0 (12.9)[b]	89.5 (6.3)[b,c]	86.0 (7.8)	0.569 (0.199)[a,c]
Multivariate Locations, mRMR					
2	$Wrist_{ZC}$, Toe_{ZC}	72.7 (21.6)[a]	85.8 (8.6)[b]	85.9 (8.4)	0.549 (0.206)[a]
3	$Wrist_{ZC}$, Toe_{ZC}, $Finger_{DI}$	73.4 (19.9)[b]	85.9 (10.0)[b]	86.9 (7.7)	0.574 (0.196)[a,c]
⩽5	$Wrist_{ZC}$, Toe_{ZC}, $Finger_{DI}$, Toe_{DI}, $Wrist_{TAT}$	76.1 (13.3)[b]	88.2 (6.2)[b]	86.8 (7.4)	0.566 (0.195)[a,c]
Heuristic removal of restless sleep					
5	$Wrist_{ZC}$, Toe_{ZC}, Toe_{TAT}, $Ankle_{DI}$, Toe_{MAXACT}	82.1 (12.5)[b,c]	90.4 (5.6)[b,c]	88.0 (6.1)[a,c]	0.611 (0.198)[a,c]

a $p < 0.05$, greater than uni-axial wrist placement.
b $p < 0.01$, greater than uni-axial wrist placement.
c $p < 0.05$, greater than single tri-axial wrist placement.
Note: Values are shown as mean (±SD).

Table 2. Class discriminibality and predictive performance of the Actiwatch and the CMAS uni-axial and tri-axial wrist placements and the heuristic feature set.

Selected features	Sp [%] (at 85% Se)	AUC [%]	Agreement [%]	κ
Actiwatch	72.8 (25.3)	86.1 (8.3)	85.6 (7.1)	0.506 (0.213)
Uni-axial, $Wrist_{ZC}$	53.0 (21.8)	79.5 (8.0)	86.8 (7.0)	0.496 (0.156)
Tri-axial, $Wrist_{ZC}$	73.4 (10.8)	84.7 (6.0)	85.2 (7.4)	0.503 (0.213)
Heuristic, $Wrist_{ZC}$, Toe_{ZC}, Toe_{TAT}, $Ankle_{DI}$, Toe_{MAXACT}	85.4 (10.6)	90.6 (6.2)	87.2 (7.5)	0.600 (0.230)

Note: Commercial actigraphy system: Actiwatch Mini, CamNTech, using the zero-crossing mode.
Due to technical issues with the Actiwatch Mini, the values summarised here are derived for each method using 13 subjects (6–13 years, median 7 years, 11 male).
Values are shown as mean (±SD)

the methods in table 1 and for the commercial actigraph and CMAS in table 2. Tri-axial accelerometry significantly improved the performance when compared to uni-axial accelerometry (specificity at 85% sensitivity: 71.3(14.2)% versus 55.2(24.7)%, $p < 0.01$). Combining locations significantly improved the ranking performance and agreement when compared to the single wrist placement (AUC for five features: 89.5(6.3)% versus 83.3(7.5)%, $p < 0.05$; Kappa agreement: 0.569(0.199) versus 0.495(0.183), $p < 0.05$). This trend in improvement was consistent across all metrics. Applying the heuristic further improved the wake detection when compared to the single wrist placement (specificity at 85% sensitivity: 82.1(12.5)% versus 71.3(14.2)%, $p < 0.05$; Kappa agreement: 0.611(0.198) versus 0.495(0.183), $p < 0.05$). The performance metrics are graphically shown in figure 4. The feature selection algorithms

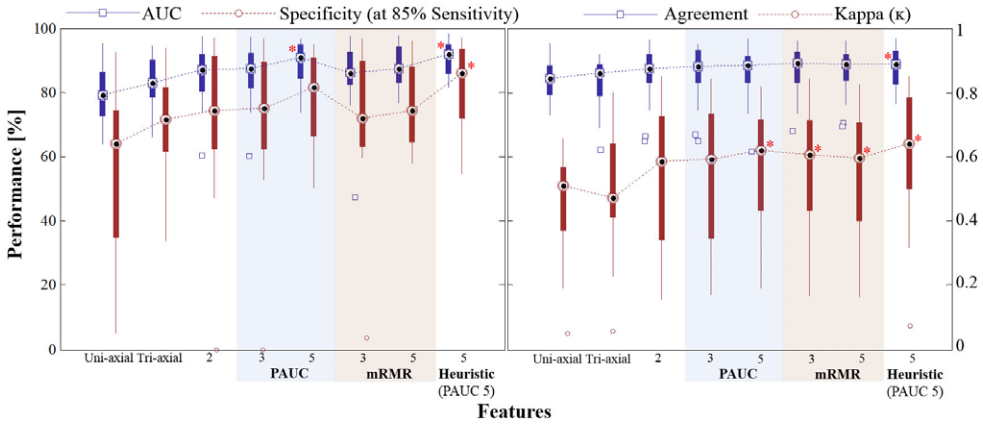

Figure 4. Boxplots of the area under curve (AUC), specificity at 85% sensitivity, agreement and Kappa agreement with the polysomnogram for the single uni-axial and tri-axial wrist accelerometer placements and the selected features from the different methods: forward feature search optimising for partial AUC (PAUC) and minimum redundancy maximum relevancy (mRMR). Also shown is the performance when applying a heuristic to remove restless sleep prior to generating the features. Significant improvements when compared to the single tri-axial wrist placement are shown as *($p < 0.05$). Shaded regions are used here to separate the feature selection techniques.

(mRMR and forward feature selection search) selected similar features and yielded consistent performance metrics. We are therefore confident that these results are not an artefact of a specific feature selection algorithm.

4. Discussion

The objective of our study was to determine whether multisite tri-axial accelerometry improved sleep and wake classification in children relative to conventional accelerometry of the wrist. We comprehensively assessed the ability of combinations of tri-axial data at the left index fingertip, left wrist, upper thorax, left ankle and left great toe to detect sleep and wakefulness on a 30 s basis using two different methods. We also applied a heuristic to remove restless sleep as a pre-processing step on the raw accelerometry data. We identify that tri-axial accelerometry measured at the wrist provides significant performance benefits beyond uni-axial accelerometry and that adopting multisite accelerometry further improves the wake detection.

4.1. Impact of multisite tri-axial accelerometry

Our results show that tri-axial accelerometry performed better than uni-axial accelerometry. The nature of this improvement is particularly well demonstrated by comparative ROC curves (figure 5(*a*)). This improvement can be mechanistically explained by the ability of tri-axial accelerometers to detect movements in all directions, the consequence of which is greater sensitivity for detecting movements. These results support existing commentary, suggesting that tri-axial data may be more effective at detecting sedentary activities (Jean-Louis *et al* 2001). We also note that the wrist was the best single accelerometer placement, as consistent with clinical practice (Meltzer *et al* 2012) and prior literature (Van Hilten *et al* 1993, Middelkoop *et al* 1997). Comparing the commercial uni-axial actigraph with the uni-axial data from the

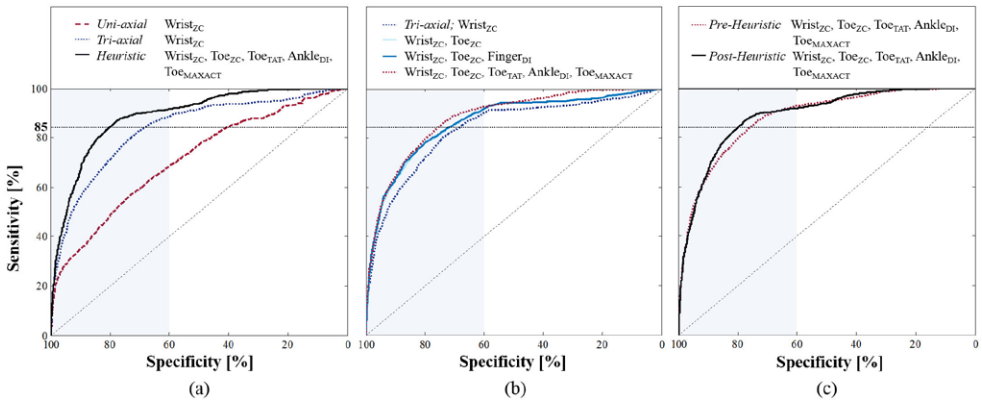

Figure 5. Example receiver operating characteristic (ROC) curves for the full population for: (*a*) the uni-axial and tri-axial wrist placement and the optimal five features, post-heuristic; (*b*) the optimal selected features (one, two, three and up to five) for PAUC; and (*c*) five features pre- and post-heuristic. The shaded area represents the portion of the ROC curves restricted to a specificity greater than 60% . The ranking performance is greatly improved when moving from uni-axial to tri-axial accelerometry and from single site accelerometry to up to five locations (in (*a*)). The heuristic further improves the performance (in (*c*)).

wrist accelerometer measured in a subset of patients confirmed that our uni-axial results are similar to the commercial system. While we have provided a comparison of the commercial actigraph with some features in this analysis, we note that care should be taken as the sample size is small (summarised in table 2).

Combining accelerometry information from multiple locations showed a trending performance improvement, which was significant for the sleep and wake discrimination ability and Kappa agreement (an example of comparative ROC curves illustrated in figure 5(*b*)). This suggests that combining multiple locations does indeed provide a greater sensitivity for detecting movements associated with sleep. Coupled with our proposed heuristic of removing movement specific to restless sleep, the wake detection and sleep and wake discrimination ability was further improved when compared to wrist accelerometry. This significant improvement in performance demonstrates that we are able to successfully isolate movements during restless sleep and differentiate those from movements associated with wakefulness. The consequences of this is an improved classification performance (as demonstrated in figure 2). Further, applying the heuristic improved all performance metrics when compared to the features prior to applying the heuristic (as shown when comparing the five features for PAUC in figure 4). The heuristic illustrates a key limitation of the conventional windowed activity counts. That is, activity counts are unable to distinguish restless sleep from active wake as activity counts only summarise the amount of activity during a fixed time period. Consequently, activity counts cannot determine what type of activity occurred. However, our proposed heuristic has demonstrated that this information is likely to aid in improving the poor specificity of conventional actigraphy.

4.2. Significance of selected features and locations

There were some interesting relationships found during our analysis. In general, the wrist was frequently selected first in the feature selection experiments, suggesting that the wrist has the

greatest sensitivity for detecting movements during wake. While we observed some improvement when combining additional accelerometer placements, the improvement became significant, particularly for specificity, when we also applied the heuristic. This suggests that the conventional activity counts are unable to fully exploit multisite data. When analysing the relationships in the features, we found that in general, features that summarised total movement within an epoch (TAT, DI, SUMPST) were complimentary to those that summarised abrupt movements (ZC, MAXACT) and so were often selected together. Conversely, features that summarised similar movements at the same locations were seldom selected together. Despite being highly relevant to the class label, there was considerable redundancy found in the wrist and finger movement. This relationship suggests that the wrist and finger accelerometers may be interchangeable. We also found this to be the case with the ankle and toe accelerometers. Interestingly, the upper thorax was seldom selected in our search, indicating that the trunk offers little additional useful information. Indeed, this agrees with previous literature which has shown that 90% of variation in trunk movement is summarised by ankle (or toe) movement (Middelkoop *et al* 1997), which is often selected in our search.

4.3. Methodological considerations

There are a number of limitations of this study. Firstly, our study population consisted of 24 children (16 male) attending the sleep laboratory for a diagnostic polysomnogram to investigate sleep-disordered breathing. This relatively small sample size, the broad range of ages (6–15 years) and the bias towards males means that care must be taken in generalising results to other cohorts. Second, the cross-validation of the same data was used in the feature selection routine and then again for training the final classifier. While this will not impact the selected features, or the relative performance between classifiers, this does introduce a bias which may positively bias the classifier performance metrics. Therefore, this impacts how these results can be interpreted in the context of other literature. While a nested cross-validation routine would alleviate this bias, computational practicality severely limits the number of features that can be considered in such an analysis. We used the Actiwatch Mini (CamNTech) as the comparative commercial actigraphy system. While this actigraph has since been remarketed for veterinary use, at the time of initial data collection the Actiwatch Mini was designed and marketed for paediatric use. As there have been more sophisticated tri-axial actigraphy systems released to the market since, future work will adopt a recent tri-axial commercial model for comparisons.

4.4. Implications for clinical practice and further research

Actigraphy performance was improved for this cohort when using multisite accelerometry, particularly when applying our heuristic. While this improvement was significant, multisite accelerometry does present some limitations in terms of setup complexity. This may seem unimportant for diagnostic sleep studies, where there are a multitude of sensors already affixed to the patient. However, this may restrict the usability in the home environment. Furthermore, we found that in practice, children occasionally play with the accelerometer units, not only interfering with the data, but at times inadvertently removing the units altogether. Minimizing the number of sensors would mitigate this problem. The tradeoff in setup complexity and performance benefits when using multisite accelerometry would need to be considered separately for each application. We also observed a significant improvement when moving from uni-axial to tri-axial accelerometry. As such, we can confirm that commercial devices recently

introduced to the market, which employ tri-axial accelerometry, will likely provide a performance benefit over the older uni-axial models.

There are a number of outcomes from this study that provide interesting avenues for further research. Firstly, it would be interesting to repeat the study design in a larger cohort of children to allow for age and gender stratifications and in an adult population. Second, as combining activity representations improved the predictive performance, we can infer that the conventional activity representations may not capture all movement information pertaining to effective sleep and wake detection. Further work could focus on improving activity representations with tri-axial accelerometry measured from the wrist, possibly focusing on time-frequency based representations, such as those applied in human activity monitoring (Bao and Intille 2004, Preece *et al* 2009). We noted that we were able to significantly improve the performance by detecting different movements using the raw multisite accelerometry data when applying our heuristic. Therefore, it may be beneficial to remove the constraint of fixed epochs and focus on detecting and classifying periods of activity with multisite accelerometry. As such, it may be interesting to determine and characterise movements that are specific to sleep or wake (Verhaert *et al* 2011, Domingues *et al* 2013). Therefore, movement detection and clustering techniques should be employed to determine whether there are distinct movements or positions that characterise sleep and wakefulness (Nguyen *et al* 2007). We have used a well understood classification technique for analysing the predictive performance of the features in this analysis. Exploring other classification techniques, such as neural networks or support vector machines, may further improve the performance.

5. Conclusion

Our analysis sought to determine whether multisite tri-axial accelerometry improved sleep and wake classification in children, relative to conventional uni-axial accelerometry measured at the wrist. We found a significant improvement in wake detection when using multisite accelerometry over single-site accelerometry with our heuristic. This highlights the potential of improving the predictive performance by removing the constraint of fixed epochs and consequently detecting movement from raw multisite accelerometry. Our results also confirm that when summarising movement with activity counts, the greatest improvement in actigraphy performance is likely to be achieved by adopting tri-axial accelerometry. The feature selection methods used in our study both consistently selected the wrist and toe placements, confirming that our results are invariant to feature selection algorithms.

Acknowledgments

This work was supported by the Queensland Government Smart State Grant and approved by an institutional human research ethics committee (ref. 1498C). APB is supported by an Australian Research Council Future Fellowship (FT110100623).

References

Adkins K W, Goldman S E, Fawkes D, Surdyka K, Wang L, Song Y and Malow B A 2012 A pilot study of shoulder placement for actigraphy in children *Behav. Sleep Med.* **10** 138–47

Ancoli-Israel S, Cole R, Alessi C, Chambers M, Moorcroft W and Pollak C P 2003 The role of actigraphy in the study of sleep and circadian rhythms (American Academy of Sleep Medicine Review Paper) *Sleep* **26** 342–92

Bao L and Intille S 2004 Activity recognition from user-annotated acceleration data *Pervasive Computing* 1–17

Beck S E and Marcus C L 2009 Pediatric polysomnography *Sleep Med. Clin.* **4** 393–406

Bradley A P 1997 The use of the area under the roc curve in the evaluation of machine learning algorithms *Pattern Recogn.* **30** 1145–59

Cios K J, Pedrycz W and Swiniarski R W 1998 *Data Mining and Knowledge Discovery* (New York: Springer)

Crespo C, Aboy M, Fernndez J R and Mojn A 2012 Automatic identification of activity—rest periods based on actigraphy *Med. Biol. Eng. Comput.* **50** 329–40

Dodd L E and Pepe M S 2003 Partial auc estimation and regression *Biometrics* **59** 614–23

Domingues A, Paiva T and Sanches J 2012 An actigraphy heterogeneous mixture model for sleep assessment *2012 Ann. Int. Conf. IEEE Engineering in Medicine and Biology Soc.* (*San Diego, CA, USA, 28 August–1 September* 2012) pp 2275–8

Domingues A, Paiva T and Sanches J 2013 Statistical characterization of actigraphy data for sleep/wake assessment *Int. J. Bioelectromagn.* **15** 13–9

Domingues A, Paiva T and Sanches J 2014 Sleep and wakefulness state detection in nocturnal actigraphy based on movement information *IEEE Trans. Biomed. Eng.* **61** 426–34

Feinstein A R and Cicchetti D V 1990 High agreement but low kappa: I the problems of two paradoxes *J. Clin. Epidemiol.* **43** 543–9

Giganti F, Ficca G, Gori S and Salzarulo P 2008 Body movements during night sleep and their relationship with sleep stages are further modified in very old subjects *Brain Res. Bull.* **75** 66–9

Gori S, Ficca G, Giganti F, Nasso I D, Murri L and Salzarulo P 2004 Body movements during night sleep in healthy elderly subjects and Their relationships with sleep stages *Brain Res. Bull.* **63** 393–7

Hutter M 2002 Distribution of mutual information *Adv. Neural Inform. Proc. Syst.* **1** 399–406

Jean-Louis G, Kripke D F, Mason W J, Elliott J A and Youngstedt S D 2001 Sleep estimation from wrist movement quantified by different actigraphic modalities *J. Neurosci. Methods* **105** 185–91

Kinder J R, Lee K A, Thompson H, Hicks K, Topp K and Madsen K A 2012 Validation of a hip-worn accelerometer in measuring sleep time in children *J. Pediatric Nursing* **27** 127–33

Lachenbruch P A and Goldstein M 1979 Discriminant analysis *Biometrics* **35** 69–85

Liu H and Motoda H 1998 *Feature Selection for Knowledge Discovery and Data Mining* (New York: Springer)

Maija-Riikka Steenari B and Aronen E T 2002 Actigraph placement and sleep estimation in children *Sleep* **25** 235–7

Meltzer L J, Montgomery-Downs H E, Insana S P and Walsh C M 2012 Use of actigraphy for assessment in pediatric sleep research *Sleep Med. Rev.* **16** 463–75

Middelkoop H A, Dam E M, Smildevan Den Doel D A and Dijk G 1997 45 h continuous quintuple-site actimetry: relations between trunk and limb movements and effects of circadian sleepwake rhythmicity *Psychophysiology* **34** 199–203

Mitra P, Murthy C and Pal S K 2002 Unsupervised feature selection using feature similarity *IEEE Trans. Pattern Anal. Mach. Intell.* **24** 301–12

Nguyen A, Moore D and McCowan I 2007 Unsupervised clustering of free-living human activities using ambulatory accelerometry *29th Ann. Int. Conf. IEEE Engineering in Medicine and Biology Soc.* (*Lyon, France, 23–27 August* 2007) pp 4895–8

Peng H, Long F and Ding C 2005 Feature selection based on mutual information criteria of max-dependency, max-relevance and min-redundancy *IEEE Trans. Pattern Anal. Mach. Intell.* **27** 1226–38

Preece J S, Goulermas Y J, Kenney P J L, Howard D, Meijer K and Crompton R 2009 Activity identification using body-mounted sensorsa review of classification techniques *Physiol. Meas.* **30** R1

Rapp M P, Nelson F, Oliver M, Bergstrom N and Cron S G 2010 Comparison of commonly used placement sites for activity monitoring *Biol. Res. Nursing* **11** 302–9

Roebuck A, Monasterio V, Gederi E, Osipov M, Behar J, Malhotra A, Penzel T and Clifford G 2014 A review of signals used in sleep analysis *Physiol. Meas.* **35** R1

Sadeh A 2011 The role and validity of actigraphy in sleep medicine: an update *Sleep Med. Rev.* **15** 259–67

Saeys Y, Inza I and Larrañaga P 2007 A review of feature selection techniques in bioinformatics *Bioinformatics* **23** 2507–17

Sazonov E *et al* 2004 Activity-based sleep–wake identification in infants *Physiol. Meas.* **25** 1291

Terrill P I, Mason D G and Wilson S J 2010 Development of a continuous multisite accelerometry system for studying movements during sleep *2010 Ann. Int. Conf. IEEE Engineering in Medicine and Biology Society* (*Buenos Aires, Argentina, 1–4 September* 2010) pp 6150–3

Thiemjarus S, Lo B, Laerhoven K and Yang G 2004 Feature selection for wireless sensor networks *1st International Workshop on Wearable and Implantable Body Sensor Networks* (*Imperial, London, UK, 6–7 April* 2004)

Van Hilten J, Middelkoop H, Kuiper S, Kramer C and Roos R 1993 Where to record motor activity: an evaluation of commonly used sites of placement for activity monitors *Electroencephalogr. Clin. Neurophysiol./Evoked Potentials Sec.* **89** 359–62

Verhaert V, Haex B, De Wilde T, Berckmans D, Vandekerckhove M, Verbraecken J and Sloten J V 2011 Unobtrusive assessment of motor patterns during sleep based on mattress indentation measurements *IEEE Trans. Inform. Technol. Biomed.* **15** 787–94

Yu L and Liu H 2004 Efficient feature selection via analysis of relevance and redundancy *J. Mach. Learning Res.* **5** 1205–24