Human Observer and Automatic Assessment of Movement Related Self-Efficacy in Chronic Pain: from Exercise to Functional Activity

Temitayo A. Olugbade, Nadia Bianchi-Berthouze, Nicolai Marquardt, and Amanda C. de C. Williams

Abstract—Clinicians tailor intervention in chronic pain rehabilitation to movement related self-efficacy (MRSE). This motivates us to investigate automatic MRSE estimation in this context towards the development of technology that is able to provide appropriate support in the absence of a clinician. We first explored clinical observer estimation, which showed that body movement behaviours, rather than facial expressions or engagement behaviours, were more pertinent to MRSE estimation during physical activity instances. Based on our findings, we built a system that estimates MRSE from bodily expressions and bodily muscle activity captured using wearable sensors. Our results (F1 scores of 0.95 and 0.78 in two physical exercise types) provide evidence of the feasibility of automatic MRSE estimation to support chronic pain physical rehabilitation. We further explored automatic estimation of MRSE with a reduced set of low-cost sensors to investigate the possibility of embedding such capabilities in ubiquitous wearable devices to support functional activity. Our evaluation for both exercise and functional activity resulted in F1 score of 0.79. This result suggests the possibility of (and calls for more studies on) MRSE estimation during everyday functioning in ubiquitous settings. We provide a discussion of the implication of our findings for relevant areas.

Index Terms— Affective computing, bodily expressions, bodily muscle activity, chronic pain, self-efficacy

1 Introduction

CHRONIC pain is pain that persists in the absence of tissue damage and is associated with disorder in the neural system [1]. It is a prevalent condition affecting over 19% of adults [2][3]. In this condition, pain is experienced even during harmless activities such as everyday movement [4] making physical functioning challenging for people with chronic pain. Meanwhile, physical functioning is important given that it represents valued goals, e.g. employment, family care. In addition, maintaining physical activeness may over time contribute to a reduction in the spread of chronic pain. Unhelpful appraisals such as the perception of functional activities as being the cause of pain or contributing to its exacerbation is associated with the persistence of pain that is symptomatic of chronic pain [5]. Addressing these psychological barriers to engagement in movement despite pain is at the core of chronic pain rehabilitation [6]. In this paper, we focus on movement related self-efficacy (MRSE)–a person's level of confidence that they can successfully execute the movements required to perform an activity–which is a key statistical predictor of elements of physical functioning [7]. In particular, we investigated the possibility of automatically detecting low MRSE with the long-term aim of designing affect aware technology for physical therapy coaching.

The classic paper of Bandura [8] proposed that self-efficacy influences the amount of effort and persistence that a person will put into the performance of the associated activity in the face of challenges. Bandura suggested that a person with low self-efficacy for an activity will tend to avoid that activity. The study of [7] indeed reveals a negative relationship between pain based MRSE and avoidance behaviour in everyday physical activities even after controlling for pain intensity and other variables such as chronic pain duration and age. This finding implies that intervention aimed at promoting physical functioning in people with chronic pain needs to address self-efficacy beliefs. In fact, it is known that clinical intervention in the rehabilitation of people with chronic pain addresses MRSE [9]. For example, [9] found that physiotherapists adjust the amount and type of positive feedback during therapy to a patient's level of self-efficacy with the purpose of balancing the provision of encouragement with the promotion of independence. The authors also found that physiotherapists tailor the difficulty and complexity of prescribed physical exercises to levels of self-efficacy rather than just the person's physical capabilities and pain levels.

Following the practice of physiotherapists, the ability to perceive a person's MRSE level in an activity could help physical rehabilitation technology tailor intervention in the absence of a clinician. Such personalised support

- *T.A. Olugbade is with University College (UCL) London Interaction Centre, UCL, Gower Street, London WC1E 6BT, United Kingdom. E-mail: t.olugbade@cs.ucl.ac.uk.*
- *N. Bianchi-Berthouze is with University College (UCL) London Interaction Centre, UCL, Gower Street, London WC1E 6BT, United Kingdom. E-mail: n.berthouze@ucl.ac.uk.*
- *N. Marquardt is with University College (UCL) London Interaction Centre, UCL, Gower Street, London WC1E 6BT, United Kingdom. E-mail: n.marquardt@ucl.ac.uk.*
- *A.C. de C. Williams is with Research Department of Clinical, Educational & Health Psychology, UCL, Gower Street, London WC1E 6BT, United Kingdom. E-mail: amanda.williams@ucl.ac.uk.*

has the potential to foster the person's MRSE and in turn promote engagement in adaptive behaviours and coping capabilities when confronted with barriers such as pain [8][10]. In addition, logging pain related behaviours and beliefs such as MRSE provides opportunity for long-term self-reflection on the influence of MRSE on movement behaviour in spite of pain experienced [10]. Logged improvement in MRSE could also provide evidence of progress in capability especially when physical capability gains are slow [9].

This paper aims to contribute an understanding of how physiotherapists estimate MRSE through observation, and to investigate the feasibility of automatic estimation of MRSE from body movement (including muscle activity) measurements during physical activity. We carried out three studies to investigate these questions. In Study 1, we explored how physiotherapists assess MRSE in people with chronic pain during physical exercises. Our observers rated the performances of these activities using three levels of MRSE. They also reported in short answers the cues they used in making their estimates. Using data from a full body motion capture system and high fidelity surface electromyography (sEMG) sensors gathered during these physical exercises, Study 2 investigated the feasibility of automatic estimation that mimics physiotherapist assessment. In Study 3, we further investigated the possibility of automatic MRSE estimation based on a minimal network of low-cost wearable sensors that track body movement (with muscle activity) during a physical exercise type and an associated functional activity.

Before presenting the above three studies, we first provide additional background on the relationship between MRSE and physical performance, and review previous work on human observer estimation of MRSE.

2 Background: MRSE and Movement

Relationship between MRSE and performance (e.g. in [7] as earlier mentioned) has been revealed in various contexts beyond chronic pain. For example, [11] found balance self-efficacy to be a predictor of falls, whereas mobility measures, such as time to complete, posture sway, knee extension, and hip abduction, objectively assessed during functional tests were not. In both these studies, [7] and [11], physical performance and self-efficacy were assessed using self-report.

In [12][13], where MRSE was measured with respect to specific instances of physical activity, self-report of self-efficacy was taken before the associated activity performance. In [12], performance outcome was measured in a wrestling match using win/loss scoring and a points system based on observation. Similarly, in [13], performance, measured in a race, was based on finish time and finish position from the first. Both studies found significant correlation between MRSE and subsequent performance outcome.

These studies give evidence of a correlation between MRSE and physical performance. While self-report could be used to measure a patient's MRSE during technology-assisted physical rehabilitation, a growing body of evidence in affective computing show that similar psychological measures (e.g., depression [14]) can be estimated by using behaviour-sensing technology. Such automatic estimation would allow the computation of MRSE as a person performs the associated activity, without interruption or calling attention to the negative elements of the activity, and providing relatively objective measurement. It also offers the possibility of providing personalised support during not only physical exercises but also functional activities. This is pertinent given that recent studies [9][15][16][17] have shown that it is important to integrate physical rehabilitation strategies during functioning rather than just during exercising. In this paper, we mainly consider exercises but also move a step (in Study 3) towards the context of functioning.

A question to be asked is, what cues could be used to automatically estimate MRSE? Observer estimation of a person's MRSE could provide a unique opportunity to understand what these cues could be. The only study to have considered observer estimation of MRSE levels before ours is [18]. In contrast to our more clinical interest, the authors considered child athletes performing gymnastic routines. They used observers from their research team, physical education teachers, and the subjects' peers. The observers were instructed to rate three levels of MRSE–"*1 for 'moves confidently', 2 for 'gives mixed signs, displaying signs both of confidence and lack of confidence', and 3 for 'does not move confidently'*" [18](pp. 467-468)–of individual subjects by looking at videotaped gymnastic performances of the subjects. The authors were interested in assessing the consistency of MRSE ratings made on different occasions by the same observers in addition to concordance between different observers. Thus, each observer repeated the same ratings a week later with access to notes of cues they had taken in the first round. A high rating consistency was found within and between observers, with absolute agreement, average measures intraclass correlation ICC = 0.97 [19]. The authors also analysed observer reported cues used in the ratings and found that observers paid particular attention to: 1) movements performed to start a routine, such as moving to the appropriate start position or fidgeting, 2) how a routine was performed, 3) timing and sequencing of phases of a routine, and 4) engagement behaviours of the subjects, such as looking excessively at the experimenter.

Despite the fact that the work of [18] was not in the context of rehabilitation, it suggests that behaviour cues can be used to infer MRSE. Hence, in our first study, we investigate how physiotherapists infer MRSE in people with chronic pain during physical activity to understand how it could be automatically estimated by technology in this context.

3 Study 1 - Observer Annotation of Movement Related Self-Efficacy

In this study, we collected observer ratings of MRSE for recorded physical exercise performances of people with chronic pain and healthy control subjects to be used as ground truth for building the proposed automatic MRSE assessment system. We also collected cues used by the observers in their ratings to inform the features to be used by the system. In this section, we describe our methods and results, and provide a discussion of our findings.

3.1 Method

3.1.1 Dataset

We used an available dataset from the EmoPain corpus [20]. The corpus consists of video, inertia based motion capture, and sEMG muscle activity data of people with chronic low back pain and healthy control participants. These data were captured while the participants performed physical exercises that are representative of everyday physical activities. The corpus does not provide MRSE labels, making the annotation study necessary to enable the use of the corpus to build the proposed automatic MRSE detection system.

For the annotation study, we used 421 video clips, 48.2% were of 17 people with chronic pain and 51.8% of 21 healthy control participants, for which the recorded persons had given consent. Each clip showed either a person with chronic pain or a healthy person, both of whom we will refer to as ‘subject’, performing Sit-to-Stand, Forward Trunk Flexion or Full Trunk Flexion. Each of these is typical of everyday functioning and is challenging for people with low back pain [21][22]. In the Sit-to-Stand, the subject stands up from seated position. The exercise was performed at two difficulty levels: a subject had to stand up at his/her own pace (less difficult) or stand up at the prompt of the instructor (more difficult). In the Forward Trunk Flexion, the subjects were instructed to reach as far forward with their trunk as they could while standing, with their arms horizontal in front of them. In the more difficult version, the participant also had to hold a 2-kilogram dumbbell. The two (objective) levels of exercise difficulty for Sit-to-Stand and Forward Trunk Flexion were only included in this study to maximize the variety in self-efficacy levels expressed in the exercise instances. For people with chronic pain, the challenge that an everyday movement poses is dependent on their emotional response to the movement, based on previous, current, or expected pain experience, rather than (solely) on these difficulty levels [5][8]. For Full Trunk Flexion, the subject reached down towards his/her toes from a standing position, similar to reaching down to pick an object up from the floor. This activity had only one difficulty level.

Each of the video clips used in the study showed the ventrolateral view of the full body of a single subject performing one of the activities (in one difficulty level).

3.1.2 Observers

We used physiotherapists as observers (hereafter referred to as ‘raters’) as they are a key part of clinical teams that support chronic pain rehabilitation and are experienced in reading movement behaviour. We recruited 30 UK physiotherapists whose number of years of physical therapy experience ranged between 1 and 36 years (median = 11.5, interquartile range = 12.75) and with pain management experience ranging from less than 1 to 32 years (median = 5, interquartile range = 6.75). We used a special design to assign the video clips in the dataset to the raters as having each rater rate all of them would have been too burdensome (about five hours in all). Each of the raters was assigned an hour rating worth of clips randomly selected such that each of the clips was rated by 4 raters.

3.1.3 Procedure

All raters had to complete a pre-annotation questionnaire where they provided information about their professional background. They also scored their own levels of confidence (hereafter referred to as ‘rater confidence’) in evaluating a patient’s MSRE. In all our instructions to the raters, we referred to MRSE as ‘movement confidence’; this was done to foster consistency between observers, similar to providing a definition for the term [20][23].

After this preliminary profiling of the raters, they were asked to look at the set of videos assigned to them and complete the MRSE ratings for each of the videos. All the videos were shown mute to compel the raters to use visual cues alone. Similar to [18], for each video clip, raters rated their estimation of the subject’s self-efficacy for the performed instance of physical activity as low, medium or high movement confidence.

In clinical practice, physiotherapists judge MRSE levels while being physically present with a patient. Given the different observation setting in this study, we were interested in understanding if being restricted to observation via video affected the ability of the raters to estimate MRSE. Thus, for comparison with the initial self-reports, rater confidence was again obtained from the raters after they completed the full annotation. To complement this, we also asked the raters how difficult they found the rating of MRSE. Finally, the raters were asked to report the nonverbal cues that informed their judgements of MRSE.

3.2 Results

3.2.1 Interrater Agreement

We were interested in understanding absolute agreement between the raters and how much the ratings could be generalised to one typical rater. This was assessed using a one-way random, absolute agreement, single measures ICC. The ICC is a standard method in designs where different sets of raters rated each observation [24].

ICC for Forward Trunk Flexion was in the poor range, ICC = 0.37 [19]; for Sit-to-Stand, it was in the fair range, ICC = 0.52; and for Full Trunk Flexion, it was also in the fair range, ICC = 0.55. Further analysis showed that rating ties accounted for a large proportion of the lack of consensus with the most frequent tie occuring between the medium and high MRSE levels.

3.2.2 Cues

We analysed the nonverbal cues that the raters reported using to assess MRSE so as to inform feature extraction for automatic estimation as well as to identify themes within the cues that may have implications for the design of an automatic observer. 29 of the raters provided 88 cues–one of the raters did not complete the post-annotation self-report. These cues were reduced to a distinct set (see Table 1) by removing both literal and seman-

TABLE 1
MRSE CUES REPORTED BY RATERS (DUPLICATES REMOVED)

	Cue Extracts		Cue Extracts
1	speed	20	unusual pattern
2	speed of starting	21	unnatural poses
3	no hesitation	22	finish position
4	hesitation	23	final posture (if it looked natural to the subject, i.e. if looking comfortable)
5	hesitation on initiating	24	length of time pose held
6	smoothness of movement	25	splinting behaviours
7	general look of relaxation	26	that they or several participants repeated movements
8	weight transfer	27	interaction or not with research team
9	ease in which they performed task	28	jerky
10	amplitude of range of movement	29	compensatory movements
11	amount of movement from trunk	30	facial expression
12	guarded behaviour	31	blink rate
13	balance saving reactions	32	how present they looked whilst doing it
14	willingness to move	33	avertal gaze
15	through sequencing	34	looking down at assesors, equipment
16	balance and alignment	35	looking down at themselves - reassurance
17	quality of movement	36	if they scanned environment around them
18	global efficiency	37	start
19	symmetry	38	observing initiation of movement

TABLE 2
THEMES THAT EMERGED FROM THE MRSE CUES REPORTED BY THE RATERS

Cue Themes	Number of Table 1 Cues	Cue Identifier in Table 1
Expression Modalities of Cues		
Body	28	1-21, 23-29
Face	5	7, 27, 30-32
Head and/or eyes	5	27, 33-36
Behaviour Elements of Cues		
Movement behaviour		
movement preparation	2	4-5
movement performance	20	1-3, 6-13, 16-19, 24-26, 28-29
movement conclusion	2	22-23
Engagement behaviour	5	27, 33-36
Facial expressions	1	30

tic duplicates. We then applied thematic analysis [25] to identify patterns in this set. Two salient themes emerged from our analysis of the final set of cues (see Table 2). These themes are discussed in the rest of the sub-section. We will present extracts from the rater reports in double quotation marks followed by our identification number for the respective rater, written in the format 'R#' where # is a number between 2 and 30, in parenthesis.

Expression Modalities of the Cues

This category highlighted the fact that physiotherapists used cues from different modalities (body, face, head, eyes) to assess a person's MRSE, although a few cues (e.g. "start" (R10)) could not be tied to any specific modality. Tables 1 and 2 show the variety of cues used by the physiotherapists. Bodily cues were the most frequently used whereas only a few raters used facial cues and cues from a combination of the head and eyes. It follows intuition that body is an important modality of MRSE. In addition, the value of bodily expressions (in comparison to other modalities) in affect detection is well established in affect studies [26][27]. It is understood that observers tend to assess bodily expressions (more than facial expressions) when the actions (or the readiness to respond to an affective experience), rather than just the mental state, of a subject are to be judged [26]. Furthermore, subjects are predisposed to bodily expressions when they are challenged with feared stimuli as they are motivated towards behaviour that avoids or mitigates perceived harm. In particular, [28] found that people with chronic pain were more likely to express bodily pain behaviours when faced with challenging physical activity. This is in contrast to facial and verbal expressions, which were more used to communicate pain to an empathetic third party.

Elements of Behaviour specified by The Cues

Beyond the modality that the cues referred to, patterns were also found to form according to elements of the behaviour that the cues specified. Three main elements of behaviour were described by the reported cues. The majority were *movement behaviour* cues as can be seen in Table 2. The movement behaviour cues extend through all the parts of the movement: '*in preparation for movement*', '*during movement performance*', and '*at the conclusion of movement*'. However, most of the cues used referred to '*during movement performance*'. Another aspect of behaviour that emerged is the *engagement behaviour* of the subjects. Raters noted whether the subject looked around the environment, at themselves, or towards or away from the experimenter. The value of the engagement behaviour as a cue of MRSE corroborates the finding of [18]. Finally, *facial expressions* were also reported, although only one rater elaborated on the specific facial expression they used: "… grimace" (R10).

3.2.3 Rater Confidence

29 of the 30 raters provided rater confidence scores before and after completing ratings of MRSE. As can be seen in Table 3, the majority of the raters had high confidence in their ratings even after completing the annotation, with

the majority reporting low levels of difficulty in completing it. In fact, although a Wilcoxon Signed Rank Test revealed a statistically significant reduction in rater confidence after the annotation task (z=-2.461, p<0.05, with a medium effect size r=-0.32), the median (= 4) remained the same. As can be expected, using Spearman Rank Order Correlation, we found a strong, negative correlation between rater confidence post-annotation and levels of rating difficulty, rho=-0.755, n=29, p<0.0001, with high levels of rating difficulty associated with lower rater confidence. We also found significant correlation between the difference in rater confidence before and after the annotation and the rating difficulty scores, rho=-0.451, n=29, p<0.05. This suggests that the slight drop in rater confidence may be associated with the perceived level of difficulty of the rating task. The perception of difficulty may be related to the difference between typical clinical settings, where MRSE levels are judged while being physically present with a patient, and the annotation settings, as mentioned in Section 3.1.3.

3.3 Discussion

An important finding from this study was that although physiotherapists seem to use cues from a combination of visual modalities to judge MRSE while observing subjects during physical activity, they rely more on body cues. As suggested by [29], bodily cues were either a modulation of movement, such as speed and smoothness, or auxiliary to movement, e.g. "general look of relaxation" (R3), "interaction or not with the research team" (R25). We found majority of the cues to be modulations of movement. As discussed earlier, the significance of bodily cues may result from their place as part of adaptive actions (such as in "splinting behaviours" (R25), "compensatory movements" (R28)) intended to protect from pain [26][28].

Despite physiotherapists not having complete agreement in rating MRSE, the agreement was no worse than for expert observers in similar affective computing investigations [20][23][30] especially where naturalistic scenarios are considered. Indeed, when comparing the agreement observed in our study with the one reported in [18], we should consider that the physical activities performed by our subjects were not part of a (sport) choreography, making the evaluation more subjective. In addition, unlike our study, the instances annotated in [18] had been pre-selected based on initial rating by the researchers although it is not clear what this entailed. It should also be noted that we used the single measures ICC which is stricter [31] than the average measures used by [18].

Finally, we found that disagreement in MRSE ratings was most likely between medium and high levels. The fewer ties with the low MRSE level may indicate that low level MRSE is characterised by more precise cues (e.g. cues 4-5, 12-13, 20-21, 25, 28-29, 33, and 35 in Table 1), which would facilitate reliable identification of the level. In contrast, other cues such as cues 1-2, 6, 9-10 require comparison to a (subjective) definition of a baseline or standard.

Our findings support body movement as the modality for our subsequent investigation of automatic MRSE level detection. In our study, we particularly focus on cues in the '*during movement performance*' phase. This initial choice is further justified by the fact that, unlike in exercise settings where physical activities are segmented (i.e. with clear beginning and ending), in everyday contexts, the margins between consecutive activities are vague. We describe the developed automatic MRSE level detection system in the rest of the paper; but first, we review literature on affect detection in the context of pain.

TABLE 3
RATER CONFIDENCE AND RATING DIFFICULTY

Scale		Number of Raters: Rater Confidence, Pre-Annotation	Rater Confidence, Post-Annotation	Rating Difficulty
not at all	0	0	0	0
	1	0	1	7
	2	0	2	10
	3	2	8	7
	4	16	10	3
	5	10	8	1
completely	6	1	0	1

4 Background: Automatic Detection of Pain Related Affect from Movement

Despite the face still receiving more attention than other affective channels in building affect aware systems (e.g., [32][33][34]), the emergence of low-cost movement sensing technology is increasingly leading researchers to target movement as an affective modality, including in clinical contexts. Beyond work aimed at assessing affective states during sedentary clinical settings [35], there is also a growing interest in their assessment during physical activity and in situ [36]. Still, to our knowledge, only [37] has attempted to develop a model for automatic MRSE estimation. Their work aims to quantify MRSE in the elderly during exercises in the home. However, their approach assumes that MRSE can be mathematically calculated from quantifications of Bandura's main self-efficacy factors [8]: performance outcome of the self, performance outcome of an observed peer, intervention, and affective experience. Unfortunately, there is no empirical evidence to support their calculation, and there was no attempt to verify that their model does indeed relate to MRSE.

More relevant to our work, [20][38] examined the automatic detection of guarding behaviour in exercise performances of people with chronic pain. In their study, guarding behaviour was rated by clinical observers. Labels derived from these ratings were used with both observable bodily expression features, consisting of ranges of full body joint angles and joint energies, and bodily muscle activity information to investigate the automatic detection functionality. They achieved an average F1 score of 0.76 for guarding [38], in contrast to 'not guarding', in two movement types. Although their work addresses the detection of pain behaviours that may correlate with levels of MRSE, they focused on a specific set of pain behaviours namely: stiffness in movement, bracing, and rubbing [20], but not MRSE. While a relation may exist between these behaviours and MRSE, the behaviours are a result of a variety of factors (including pain and related fear [5]).

A growing body of work has aimed to automatically infer self-reported pain levels rather than pain related observed behaviour (e.g. guarding). Among the most recent works are the studies of [39] and [40][41]. Using, motion capture sensors mounted on screws inserted in the spine during surgery, [39] automatically classified 11 points of pain intensities of people with chronic pain with maximum error of 0.25 points. Using less invasive wearable motion capture and muscle sensors, [40][41] classified two levels of pain within the chronic pain group in addition to a healthy control group with a mean accuracy of 0.87 over three sets of physical exercise types. Though these works do not investigate MRSE, they provide sufficient evidence of the informative power of body movement features for assessing subjective experiences.

In the following sections, we describe two studies we carried out to assess the feasibility of automatic MRSE level detection from bodily cues, including muscle activity features. In Study 2, our investigation is based on physical exercise performances, using a subset of the EmoPain dataset used in Study 1. We extended this in Study 3 by investigating the feasibility of automatic MRSE level using a new corpus of data (named Ubi-EmoPain) acquired with a reduced set of sensors and including functional activity.

5 Study 2 - Automatic Detection of MRSE Levels

Following from the annotations obtained in Study 1 and the understanding of MRSE cues gained from the study, we investigated the automatic detection of levels of self-efficacy for physical exercises instances, from body movement behaviour during their performances. This study is based on data acquired using commercial full body motion capture (comprising 18 sensors) and (set of 4) sEMG sensors.

5.1 Method

5.1.1 Dataset

We used data from the EmoPain corpus [20] labelled with the MRSE ratings obtained in Study 1 for automatic detection using the instances without rating ties and instances for which the four raters were not split between all 3 MRSE levels. This resulted in a subset of 287 instances in total for the 3 exercises.

We recomputed the ICC to verify the reliability of the MRSE ratings in this subset. For Sit-to-Stand and Forward Trunk Flexion, the agreement values improved from poor and fair ranges to the good range, ICC = 0.67 and 0.66 respectively. It improved from the fair range to the excellent range for Full Trunk Flexion, ICC = 0.84 [19].

To define the ground truth, for each instance, we computed the mode of the physiotherapist ratings (note that the modal and the median ratings were the same for all the instances).

We used the motion capture and sEMG data of the EmoPain corpus, corresponding to the aforementioned ratings to derive features for automatic detection. The EmoPain motion capture data [20] comprises three-dimensional positions of 26 full body joints (see Table 4's Joints Map). These were captured at 60Hz using the Animazoo IGS190, based on a network of 18 gyroscopes. The BTS FreeEMG300 wireless sEMG was used to track muscle activity at 1000Hz. The muscle activity data are upper envelopes of rectified sEMG profiles (with baseline of 0 for each subject) captured bilaterally from the trapezius and L4/5 lumbar paraspinal muscles (Table 4's Muscles Map) [20].

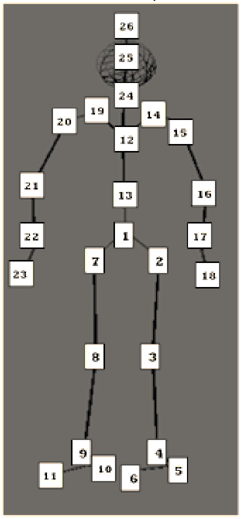
As a result of unavailability of motion capture and/or sEMG data due to unexpected disconnection in the data capture software during data capture, some instances had to be excluded. This led to a dataset of 20 instances of Forward Trunk Flexion from 7 chronic low back pain and 9 healthy participants, and 63 instances of Sit-to-Stand from 11 chronic low back pain and 8 healthy participants. There were only six instances of Full Trunk Flexion so the instances for this movement type were not included in our analysis.

5.1.2 Feature Extraction

A total of 29 features (per exercise instance) were extracted from the motion capture and sEMG data, based on the cues used by physiotherapists in Study 1. A list of the features and our computation of each are given in Table 4.

Speed was the most frequently reported cue and we extracted the mean speed of the hands, legs, shoulder, and trunk to characterise it. Speed was computed as the Euclidean distance between succeeding positions in a three-dimensional coordinate system. Fluidity or smoothness was the second most frequently reported cue and so we also extracted this feature. We derived the fluidity of a movement as a vector of the individual smoothness indices of joints involved in the movement [42]. The smoothness index of a joint during a movement was computed as the spectral arc length of the movement speed profile of the joint during the movement [43]. This was computed for hand, legs, shoulder, and trunk joints. Range of movement was another cue important to the observers. To characterise this, we used 13 features representing ranges of full body joint angles. The computation of these angles came from the work of [20]: each angle was computed as the acute angle of a joint with respect to two other joints. These angles were calculated for the pelvic, knee, neck, elbow, and collarbone joints. Since "guarded behavior" was another cue that emerged from Study 1, we used features proposed in [20] for the automatic detection of guarding, however we did not classify guarding per se. In addition to the range of joint angles earlier mentioned, the authors used joint energies and mean muscle activity as features. Thus, we extracted the sum of joint energies as the sum of the squares of angular velocity for each of the 13 joints with range of angles extracted, based on the algorithm of [44].

TABLE 4
AUTOMATIC DETECTION FEATURES

ID	Features	Formulae	Maps
S-	Joint speed	$\frac{\sum_{t=2}^{T} s_t}{T}$ $\quad s = \left[\sqrt{\left(f_{x_t}-f_{x_{t-1}}\right)^2+\left(f_{y_t}-f_{y_{t-1}}\right)^2+\left(f_{z_t}-f_{z_{t-1}}\right)^2}, \forall t\right]$	Joints Map 
1	hands (mean of left and right)	left hand *f* = {17}, right hand *f* = {22}	
2	lower legs (mean of left and right)	left lower leg *f* = {3}, right lower leg *f* = {8}	
3	upper legs (mean of left and right)	left upper leg *f* = {2}, right upper leg *f* = {7}	
4	shoulder	*f* = {14}	
5	trunk	*f* = {13}	
R-	Joint angles range (*wrt ≡ with respect to*)	$\max g - \min g \quad g = \left[\tan^{-1}\left(\frac{\left\Vert a_t \times b_t\right\Vert}{a_t \bullet b_t}\right), \forall t\right]$	
6	pelvic wrt head and left foot	*a* = {26}-{1}, *b* = {4}-{1}	
7	pelvic wrt head and right foot	*a* = {26}-{1}, *b* = {9}-{1}	
8	pelvic wrt trunk and left knee	*a* = {12}-{1}, *b* = {3}-{1}	
9	pelvic wrt trunk and right knee	*a* = {12}-{1}, *b* = {8}-{1}	
10	left knee	*a* = {2}-{3}, *b* = {4}-{3}	
11	right knee	*a* = {7}-{8}, *b* = {9}-{8}	
12	left elbow	*a* = {15}-{16}, *b* = {17}-{16}	
13	right elbow	*a* = {20}-{21}, *b* = {22}-{21}	
14	left shoulder (protraction)	*a* = {24}-{14}, *b* = {15}-{14}	
15	right shoulder (protraction)	*a* = {24}-{19}, *b* = {20}-{19}	
16	left shoulder (abduction/adduction)	*a* = {16}-{14}, *b* = {2}-{14}	
17	right shoulder (abduction/adduction)	*a* = {21}-{19}, *b* = {7}-{19}	
18	neck	*a* = {26}-{24}, *b* = {13}-{24}	
E- 19	Sum of energies	$\sum_{\forall g^*}\left(\sum_{t=2}^{T}\left(g^*_t - g^*_{t-1}\right)\right)^2 \quad g^* = smoothed\ g$	
Y- 20	Symmetry	$\frac{\sum_{t=1}^{T}\left(\left\vert \{14\}_{x_t} - \{19\}_{x_t}\right\vert\right)}{T} + \frac{\sum_{t=1}^{T}\left(\left\vert \{14\}_{z_t} - \{19\}_{z_t}\right\vert\right)}{T}$	
M-	Mean muscle activity	$\frac{\sum_{t=1}^{T} h_t}{T}$	
21	right lower	*h* = EMG1	
22	left lower	*h* = EMG2	 Muscles Map
23	right upper	*h* = EMG3	
24	left upper	*h* = EMG4	
F-	Joint fluidity [43]	$-\sum_{k=1}^{K_c-1}\sqrt{\left(\frac{1}{K_c-1}\right)^2+\left(\Delta\hat{V}_k\right)}, \quad \Delta\hat{V}_k = \hat{V}_k - \hat{V}_k - 1,\ k \in [1, K-1],$ $\hat{V}_k = \frac{V_k}{V_{k=0}},\ V_k = \left\vert \mathrm{fft}(s_{zp})\right\vert,\ k \in [0, K-1],\ s_{zp_t} = \begin{cases} s_t, & 0 \le t \le T-1 \\ 0, & T \le t \le K-1 \end{cases}$ K_C = discrete time Fourier transform index that corresponds to cut off frequency, fft = K − point fast fourier transform, $K = 2^{\mathrm{roundup}\ (\log_2 T) + 4}$	
25	hands	*s* = S-1	
26	lower legs	*s* = S-2	
27	upper legs	*s* = S-3	
28	shoulder	*s* = S-4	
29	trunk	*s* = S-5	

t = 1, …, *T* (the number of frames); *num* = 1, 2, …, 26; # = 1, 2, 3, 4
{num} = profile of the three-dimensional position of the joint labelled as *num* in the Joints Map between time *t*=1 and time *t*=*T*
$\{num\}_x$ = *x*-component of {*num*}; $\{num\}_y$ = *y*-component of {*num*}

We also extracted the mean muscle activity for each of the four tracked muscles. Finally, we characterised symmetry, another cue from Study 1. For this, we extracted the amount of dissymmetry, as the sum, for the coronal and axial planes, of the mean distance between the left and right collarbone joints along those axes. As these features used were not dependent on lengths of anatomical segments of the subjects, subject-dependent feature normalisation was not necessary.

5.1.3 Classification

The aim of the study was to automatically detect three levels of MRSE: low, medium, and high MRSE. In the study, automatic detection investigation was done separately for the two movement types (Sit-to-Stand and Forward Trunk Flexion) as we expected that discriminative MRSE features will differ between them. Although interesting, we did not consider further separation by exercise difficulty level due to the size of the dataset, and as previously mentioned, our interest is in automatic detection of MRSE levels regardless of objective difficulty levels of the movements. As classification is the typical approach used with (a small number of ordinal) discrete labels, we treated the automatic detection problem as a three-label classification problem. We used two standard classification algorithms that have shown to be effective in the discrimination of bodily expressions, e.g. in [35][38][45][46][47][48]: Random Forest (RF) and Support Vector Machine (SVM).

The RF is a multi-label classification algorithm that uses an ensemble of decision trees [49]. The RF we used in this study is a balanced Random Forest, which deals with imbalance in data by forcing oversampling of the minority classes and undersampling of the dominant classes [50]. We implemented this in MATLAB 2016a using the *Treebagger* function. We used grid search to set the hyperparameters for the RF: 50 trees for both Forward Trunk Flexion and Sit-to-Stand and one feature used to split each node for Forward Trunk Flexion but the square root of the size of the feature set for Sit-to-Stand.

The SVM, on the other hand, uses a hyperplane of maximum separation (with some allowance *C*) between two classes for classification [51]. We used a multi-layer SVM similar to [40][41] such that a first SVM, SVM1, was used to discriminate between high MRSE and lower levels of MRSE and a second SVM, SVM2, was used to further differentiate the latter as either low or medium MRSE. To deal with class imbalance with the SVM, we set the regularization parameter *C* to be rescaled such that it is higher for the minority class, forcing the hyperplane away from this class [52]. We also implemented the SVM in MATLAB 2016a, using the *svmtrain* function; grid search was used to find the optimal hyperparameters. For Forward Trunk Flexion, the optimal model was a linear polynomial SVM with $C = 0.01$ and 10 for SVM1 and SVM2 respectively. $C = 10$ was the optimum for the two SVMs of Sit-to-Stand; the optimal kernels for SVM1 and SVM2 were polynomials of degrees 1 and 3 respectively.

We used leave-one-subject-out cross-validation to evaluate the classification models to ensure generalisation capabilities to unseen subjects, as is standard in affective computing.

5.1.4 Feature Subset Selection

We were interested in optimising the feature set for two reasons: i) to maximise classification performance, and ii) to minimise the number of channels to be tracked for automatic detection. The latter will augment the feasibility of automatic MRSE level detection in ubiquitous settings as it would minimise the amount of sensors that the user needs to wear for MRSE monitoring. To achieve this, we performed feature subset selection on the feature set.

We used two approaches for selection. On one hand, we employed traditional statistics using linear mixed models, a standard method for understanding inter-variable relationships with repeated measures, with bootstrap size of about 1000. This was done using IBM SPSS Statistics 22. We also used a wrapper based selection, which is standard in machine learning. Wrapper based algorithms have the advantage of tailoring selection to the specific classification algorithm to be used [53]. In addition, they take into consideration the interaction between multiple features. The selection algorithm we used was a breadth-first tree search. To reduce the running time of the search, rather than an exhaustive search, each tree node was visited only if the node had a better value (i.e. classification performance) than its parent. This is similar to the Branch and Bound method of [54] where a node is visited only if its value is higher than a set bound based on the assumption that the successors of a node will do no better than the node. While this assumption does not always hold true, it allows faster discovery of a feature subset smaller in size than and at least as good in classification performance as the original feature set. To further reduce the running time of the algorithm, each node was also required to be of the best n=200 of its peers–other nodes in the tree whose parents have been visited–to be visited. When the additional criterion was enforced, peer nodes were visited, if they met it, in decreasing order of their values.

Feature selection was done using leave-one-subject-out cross-validation, i.e. each feature subset's superiority (over its parent) was determined by cross-validation. The cross-validation was done over the whole dataset, as the limited size of the dataset did not permit the use of a validation set completely separate from the set used for evaluation. Once the featured had been obtimized, a second cross-validation was run to evaluate the efficiency of the optimized model.

5.2 Results

In this section, we present the automatic detection performances, with each result based on aggregation of the automatic detection outputs from all folds. We also discuss the findings from the feature subset selections.

5.2.1 Classification Performance

As can be seen in Table 5, for both the RF and the SVM, classification of MRSE levels for the Forward Trunk Flexion exercise was well above chance level. Using the linear mixed model selection of features from the original set

TABLE 5
MRSE DETECTION PERFORMANCE

FORWARD TRUNK FLEXION						
	RF			SVM		
	ALL	LMM	W	ALL	LMM	W
F1 low	0.80	0.55	0.73	0.73	0.83	**1**
F1 medium	0.63	0.47	0.84	0.67	0.75	**0.95**
F1 high	0.55	0.67	0.80	0.73	0.83	**0.91**
average F1	0.66	0.56	0.79	0.71	0.80	**0.95**
accuracy	0.65	0.55	0.80	0.70	0.80	**0.95**
SIT-TO-STAND						
	RF			SVM		
	ALL	LMM	W	ALL	LMM	W
F1 low	0.60	0.89	0.73	**1**	**1**	**1**
F1 medium	0.47	0.46	0.54	0.32	0.63	**0.57**
F1 high	0.79	0.74	**0.82**	0.59	0.81	0.76
average F1	0.62	0.70	0.70	0.64	0.81	**0.78**
accuracy	0.68	0.67	**0.73**	0.52	0.76	0.71

ALL = all 29 features, LMM = linear mixed model selected features, W = wrapper based selection of features

TABLE 6
CONFUSION MATRICES USING THE WRAPPER BASED SELECTED FEATURES

FORWARD TRUNK FLEXION - SVM				
		Automatic Classification		
	MRSE LEVEL	low	medium	high
Ground Truth	low	**5** (**100**%)	0 (0%)	0 (0%)
	medium	0 (0%)	**9** (**90**%)	1 (10%)
	high	0 (0%)	0 (0%)	**5** (**100**%)

TABLE 7
CONFUSION MATRICES USING THE WRAPPER BASED SELECTED FEATURES

SIT-TO-STAND - RF				
		Automatic Classification		
	MRSE LEVEL	low	medium	high
Ground Truth	low	**4** (**100**%)	0 (0%)	0 (0%)
	medium	3 (15%)	**10** (**50**%)	7 (35%)
	high	0 (0%)	7 (18%)	**32** (**82**%)
SIT-TO-STAND - SVM				
		Automatic Classification		
	MRSE LEVEL	low	medium	high
Ground Truth	low	**4** (**100**%)	0 (0%)	0 (0%)
	medium	0 (0%)	**12** (**60**%)	8 (40%)
	high	0 (0%)	10 (26%)	**29** (**74**%)

improved the classification performance of the SVM model while it generally worsened it in the RF model. On the other hand, the wrapper based selection of features resulted in higher classification performance in both the RF and SVM models. For the RF model, classification was improved by about 20% average F1 score while it was improved by about 34% average F1 score for the SVM model. In fact, as can be seen in Table 6, for the SVM, the wrapper based search feature subset resulted in near-perfect classification where only one instance was misclassified. Unsurprisingly, as this matches the trend found in the observer estimations, the misclassification is a medium MRSE instance misclassified as high MRSE.

For Sit-to-Stand, classification performance was better than chance level except for the SVM classification of the medium level MRSE using the non-optimised feature set. This improved considerably with feature subset selection especially with the SVM. Using the wrapper based selection, for both the RF and SVM, there was perfect recall of the low class as can be seen in Table 7. In contrast, both precision and recall were poorest for the medium class with both algorithms. Majority of the misclassifications of medium MRSE was with high MRSE. This is similar to the finding with human observers in Study 1. The Sit-to-Stand is a complex movement not only because it involves coordination between several joints [55], but also due to the dependence of this coordination on contextual information particularly the height of the seat with respect to the subject's height and initial feet positions [22]. This complexity may account for the higher difficulty in automatically discriminating between medium and high MRSE levels in this movement type compared with the Forward Trunk Flexion.

Further analysis of the classification performance in both Forward Trunk Flexion and Sit-to-Stand showed agreement of the classification models with the human observers of Study 1 as ICC = 0.96 and 0.69 respectively for the two movements. This is higher than the level of agreement found between the human observers themselves. Although this comparison is unfair as technology is trained and tested on a restricted set of data provided to it, whereas a variety of experiences feed into the judgement of human observers, still it suggests that the models have good generalisation capabilities.

5.2.2 Features Analysis

In this sub-section, we report our findings on the discriminative power of the features based on optimisation using feature subset selection. We refer to the features using the numberings in Table 4; in our numbering convention, S-#, *R*-#, *E*-#, *Y*-#, *M*-#, and *F*-# refer to speed, range of joint angles, energy, symmetry, muscle activity, and fluidity features respectively.

Forward Trunk Flexion

In Forward Trunk Flexion, of the 29 features, only leg speed features *S*-2 and *S*-3, and range of trunk flexion features *R*-6 and *R*-7 were selected using the linear mixed model method; each of these features was found to significantly increase in value with increasing MRSE, $p<0.05$. It

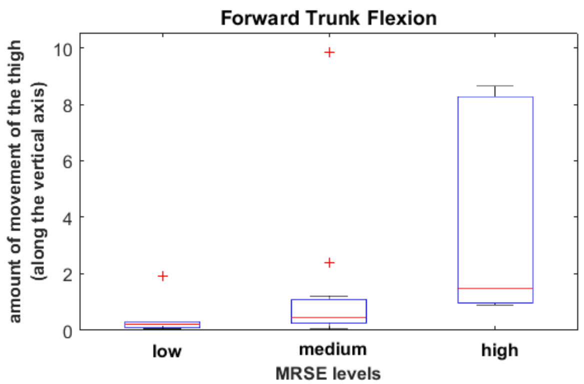


Fig. 1. Box-plots of the amount of translational movement of the thigh along the vertical axis in Forward Trunk Flexion for each MRSE level.

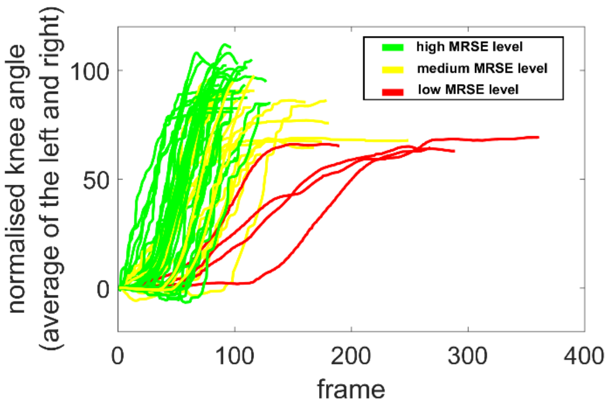


Fig. 3. Knee angle profiles (in degrees) during Sit-to-Stand instances performed with high, medium, and low MRSE.

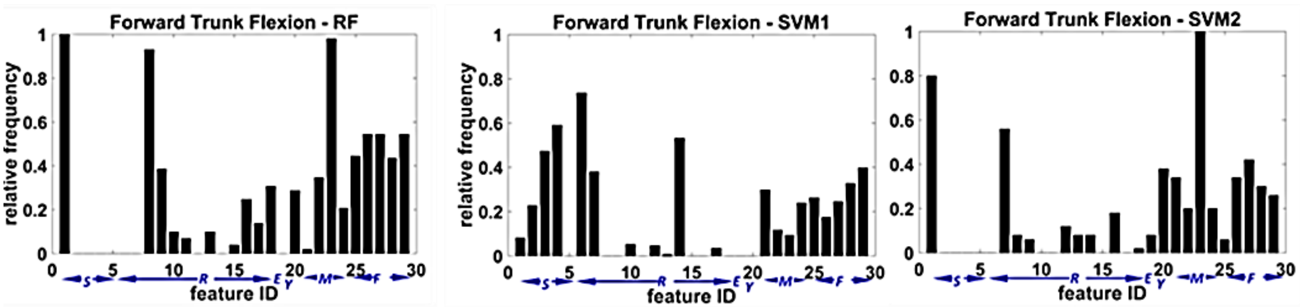


Fig. 2. Forward Trunk Flexion - Relative frequencies of the 29 features in the optimal subsets for the RF (left), SVM1 (middle), and SVM2 (right). The feature ID follows the numbering in Table 4: S, R, E, Y, M, and F = speed, range of angles, energy, symmetry, muscle activity, and fluidity features respectively.

is not surprising that the range of trunk flexion was found to be significantly related to the level of MRSE as it is the challenging element of forward trunk flexion for people with chronic low back pain [21] and was highlighted by the observers as a cue they used in their assessment of MRSE (see Table 1). The role of the speed of the upper and lower leg joints (and in essence the thighs) as a cue in Forward Trunk Flexion is on the other hand not immediately obvious. However, the revealed discriminative power of the feature suggests that smaller movement, and so low speed, of the thighs may be related to rigidity and 'guarded behaviour', which was reported as a cue in Study 1 by the physiotherapists. Indeed, Fig. 1 shows that there was relatively little translational movement of the thighs in the vertical axis for low MRSE level instances in contrast with the other levels of MRSE.

The wrapper based selection returned multiple optimal subsets within the search constraints used for each of the RF, SVM1, and SVM2. For the RF, 101 subsets were returned. Fig. 2-left shows the relative frequencies of the features in the subsets. Similarly, 171 and 50 subsets were returned for SVM1 and SVM2 respectively. Fig. 2-middle and Fig. 2-right show the relative frequencies of the 29 features in these subsets. For both the RF and the SVM, the right trapezius muscle activity feature *M*-23 seems to be a critical feature, appearing in all the subsets for RF and SVM2; the hand speed feature *S*-1 appears to be about as important.

A three-feature set was the minimal size among the optimal subsets returned for the RF. There were two such sets with the combinations: {*S*-1, *R*-14, *M*-23} and {*S*-1, *R*-8, *M*-23}. The minimal size subsets for SVM1 had two features each with the combination: {*S*-4, *R*-14}, {*R*-6, *F*-25}, and {*S*-3, *R*-6}. There were eight minimum sized subsets for SVM2 with four features each. They all included *S*-1 and *M*-23.

Sit-to-Stand

For Sit-to-Stand, range of knee motion features *R*-10 and *R*-11, energy feature *E*-19, range of elbow movement feature *R*-12, and trapezius muscle activity features *M*-24 were selected using the linear mixed model method. The first three were found to significantly increase in magnitude with increasing MRSE (p<0.01, p<0.05, p=0.001 respectively). Further analysis suggests that people with lower MRSE tend to start Sit-to-Stand with the feet as far forward as possible and so do not require as much knee flexion to complete the movement as those with higher MRSE (see Fig. 3). This may be a protective strategy aimed at limiting the amount of flexion necessary in the lumbar spine. *M*-24 increased in the order high→low→medium of MRSE level, $p<0.01$. This trend may be tied to the aforementioned finding for initial foot positioning: starting the Sit-to-Stand with the feet far forward (more likely in people with lower MRSE) demands the use of the shoulders or the use of the arms braced on the thigh or the seat edge in pushing up in seat off. This may also explain the finding of significant increase in range of elbow movement with decrease in MRSE level ($p<0.01$).

There was only one optimal feature combination found for the RF and it comprised speed, range of motion, energy, and muscle activity features, *S*-5, *R*-18, *E*-19, and *M*-22. Only one optimal subset was also returned for SVM1: *S*-2, *S*-3, *R*-7, *R*-9, *R*-14, *E*-19, *M*-23, and *F*-28, i.e. speed, range of motion, energy, muscle activity, and fluidity features. In contrast, there were 86 optimal subsets returned for SVM2; Fig. 4 shows the relative frequencies of the features in these subsets. The minimal subsets were of two features. There were 19 sets.

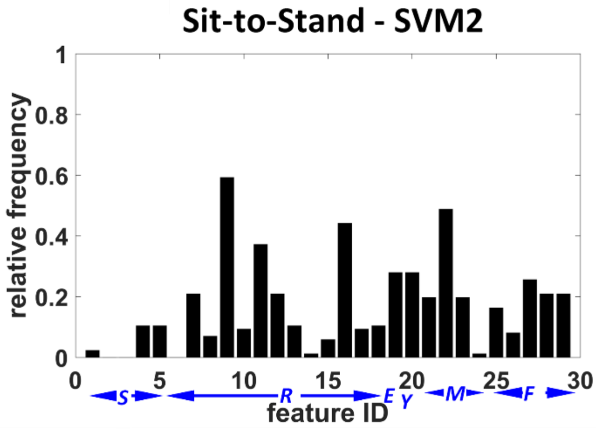


Fig. 4. Sit-to-Stand - Relative frequencies of the 29 features in the optimal subsets returned for SVM2. The feature ID follows the numbering in Table 4: S, R, E, Y, M, and F refer to speed, range of joint angles, energy, symmetry, muscle activity, and fluidity features respectively.

6 Study 3 - Automatic MRSE Level Detection Based on a Minimised Set of Low-Cost Sensors

Our long-term aim is for technology to provide ubiquitous support in physical rehabilitation, which is not just about physical exercising but rather, is centred around everyday functioning [10]. The studies of [9][15][16] show that for people with chronic pain, physical exercises are mainly integrated into everyday activity (e.g. stretching to clean a higher shelf, bending to load the washing machine). Indeed, because people with chronic pain have to manage limited physical and psychological (e.g. fear of increased pain, anxiety towards movement) resources, they tend to save such resources for everyday activities that are necessary and use these activities as their main source of exercise. In such everyday settings, it is critical to use portable, wearable sensing devices rather than a camera based system or a cumbersome wearable system (as in Study 2). It is also necessary to understand the feasibility of low-cost systems for mass deployment. Study 3 attempts a first step in this direction.

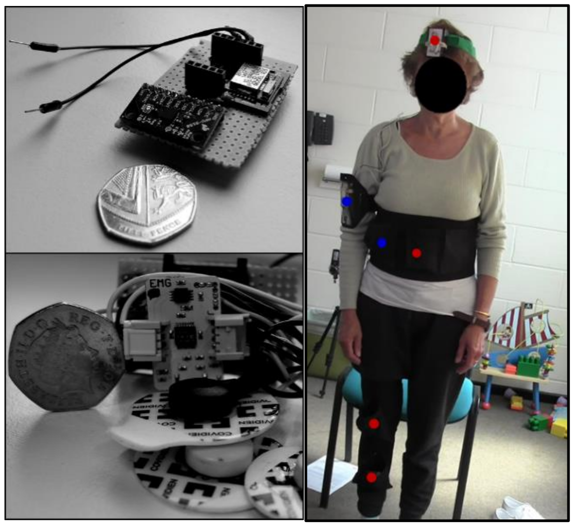


Fig. 5. Top-Left - A motion capture sensing unit (a MPU9150 sensor and a microcontroller). Bottom-Left - A sEMG sensing unit (a BITalino sEMG sensor and a microcontroller). Right: A participant wearing 4 of the motion capture units (red dots) on the head, the trunk, the right upper and lower leg, and 2 of the sEMG units (blue dots) on the arm with electrodes attached to the right trapezius muscle and on the trunk with electrodes attached to the right lumbar paraspinal muscle.

6.1 Method

6.1.1 Low-Cost Sensing Prototype

We built a custom, wearable sensing prototype that measures the orientation of anatomical segments and muscle activity. The sensors we used for our investigation were the SparkFun MPU9150 integrated triaxial accelerometer, gyroscope, and magnetometer board [56] (see Fig. 5. Top-left) and the BITalino sEMG sensor [57] (see Fig. 5. Bottom-Left). As the magnetometer is prone to magnetic interference, we used only accelerometer and gyroscope data in this study; the use of these for estimating orientation has been validated in [58][59]. We used a complementary filter [60] to estimate orientation from these data locally on an RFduino RFD22301 microcontroller. The sEMG sensor was used to estimate muscle activity within ±1.65 millivolts range; computation of this was also done locally on the microcontroller. For data collection, each of the motion capture and sEMG sensing units were fitted into box enclosures of size 27 x 50 x 75 millimetres and weight of 0.030 kilograms. Each orientation sensing unit with its battery and enclosure packaging weighed 0.052 kilograms and each sEMG unit weighed 0.069 kilograms with its leads, battery, and enclosure package. A 3.7 Volts polymer lithium ion battery was used for each sensing unit. Data from these sensing units were transmitted via Bluetooth Low Energy (BLE) to a mobile application we developed. Orientation was recorded at about 45Hz while sEMG was recorded at about 65Hz.

6.1.2 Dataset

We used the low-cost sensing prototype to acquire a new dataset (named Ubi-EmoPain) of body movement data from 12 people with chronic low back pain. The participants wore units of the prototype while performing a series of movements, including 4 instances of Forward Trunk Flexion. Participants were asked to perform each of these four instances as they normally would rather than through a prescriptive model. They were also instructed to only reach as far forward as they were able to. Two versions of this movement were requested: in one version (the Exercise) of the movement, participants were asked to reach forward starting from standing position as if to reach for a distant object without any restriction on how this should be executed. There were three repetitions of the Exercise. In the second version (the Functional), participants were asked to retrieve an empty cardboard box from the wall in front of them; the box was attached to the wall using pressure sensitive adhesive and a table between them and the wall necessitated stretching forward to reach the box. While the Exercise Forward Trunk Flexion represents a typical exercise simulating routine forward trunk flexion in everyday functioning, the Functional Forward Trunk Flexion is a functional activity–taking an object from a wall–and enables us to investigate automatic detection in 'near-wild' settings, a step closer to everyday scenarios. In [61], people with chronic pain reported being more anxious and less confident when performing this movement during functional activity rather than as exercise. For this reason, only one instance of the functional movement was requested. This is a typical problem when building datasets in sensitive real life contexts where the possibility of collecting the data is reduced due to ethical issues. The Exercise and Functional instances were not performed in immediate succession: other physical activities (e.g. Sit-to-Stand), and brief rest breaks for some participants, occurred between them. This was intentionally arranged to prevent monotonicity or lack of inter-subject variation due to repetitiveness. The Functional Forward Trunk Flexion was always performed as the second instance of the movement. Although 48 movement instances in total were completed, due to BLE interference, only 45 movement instances were successfully recorded. Thus, the analysis reported in this section is based on these 45 instances.

To record data for these movement performances, sensors were placed as shown in Fig. 5-Right. Four of the orientation sensing units were placed respectively on the head, trunk, right upper and lower legs, which are the main anatomical segments whose movements were found in Study 2 to significantly differentiate MRSE levels. The units were attached using adjustable accessories. Two sEMG units were used with each placed on the right trapezius and the right lumbar paraspinal, L4/5. The reference leads of the sEMG sensors were placed on the cervical bone and along the spine for the trapezius and lumbar paraspinal muscles respectively. Sensors were only placed on the right side for the bilateral segments (the legs and the muscles) as data transmission suffered from interference when additional sensor units were used. Although an unexpected constraint, this aligned with our aim to minimize the number of sensors necessary for automatic detection.

Similar to Study 2, the speed, sum of energy, spectral arc length of the speed profile, and range of movement of the head, trunk, and upper and lower legs were extracted as features (per exercise instance) based on the findings in Study 1. The first three features were computed in a similar way to the methods used in Study 2 (see Table 4). Two-dimensional orientation (i.e. pitch and yaw) profiles were used in place of three-dimensional position profiles. Range of movement was calculated as the range in orientation along the pitch axis. We additionally extracted the mean angular jerk for each of the segments (also per exercise instance), computed as the third derivative of orientation, to characterise fluidity in addition to the speed spectral arc length. This was the only feature used in Study 3 but not in Study 2. The feature was added to supplement the original fluidity features, which did not have as much discriminative power with respect to self-efficacy for the Forward Trunk Flexion as the speed and range of motion feature features (see Study 2). Mean muscle activity was computed (per exercise instance) from smoothed rectified sEMG data (normalized to baseline of 0 for each instance); smoothing was done by Savitzky-Golay filtering [62] after full-wave rectification.

In this study, MRSE ground truth was collected directly from the participants using a single item self-report: before completing each instance of activity, participants were asked to report their level of confidence, on a scale from 0 for not at all confident to 10 for completely confident, about being able to perform that instance of the activity. As preliminary investigation, only two levels of MRSE were considered in this study. These were obtained from the self-reports as lower level MRSE for self-report of 5 or less and higher level MRSE for self-report of more than 5/10 (in future work, more levels will be considered). This resulted in 14 instances of lower level MRSE and 31 instances of higher level MRSE.

6.1.3 Dealing with Missing Data

Of the 45 movement instances from the 12 participants, partial body movement data for 8 participants were missing due to BLE interference and inadequateness of the attachment accessories for some participants, and although each participant had at least one segment tracked, there were 23% missing features values. We explored two approaches to deal with missing values for automatic detection: imputation to recover missing feature values [63], and the use of surrogate splits (in tree based classification) with incomplete feature values [64]. As decision trees needed to be used for the approach based on surrogate splits, for the sake of comparison, we used decision trees for classification with both approaches. The decision tree was implemented using MATLAB 2016a function *fitctree* with no surrogates and all surrogates for the imputed and non-imputed dataset respectively. We also explored classification based on the SVM and the RF with imputation.

Our dataset was not normally distributed, and so we

TABLE 8
F1 SCORES AND ACCURACY FOR THE INCOMPLETE DATASET AND FOR THE PRIMARY IMPUTED DATASET (USING THE DECISION TREE) WITHOUT AND WITH FEATURE SET OPTIMISATION

	Incomplete Dataset	Imputed (no optimisation)	Imputed (optimisation)
F1 lower MRSE	0.52	0.54	**0.71**
F1 higher MRSE	0.75	0.81	**0.87**
average F1	0.64	0.68	**0.79**
accuracy	0.67	0.73	**0.82**

TABLE 9
F1 SCORES AND ACCURACY FOR THE PRIMARY IMPUTED DATASET USING THE SVM AND THE RF (WITHOUT FEATURE SET OPTIMISATION)

	SVM	RF
F1 lower MRSE	0.45	0.36
F1 higher MRSE	0.71	0.71
average F1	0.58	0.54
accuracy	0.62	0.60

TABLE 10
F1 SCORES AND ACCURACY FOR *M* SECONDARY IMPUTED DATASETS

M	1	2	3	4	5
F1 lower MRSE	0.74	0.60	0.48	0.50	0.54
F1 higher MRSE	0.89	0.80	0.80	0.72	0.81
average F1	0.82	0.70	0.64	0.61	0.675
accuracy	0.84	0.73	0.71	0.64	0.73

could not use the standard expectation-maximization for imputation [63]. We instead used linear regression for imputation; this was done with IBM SPSS Statistics 22. To build the regression model used for imputation, we included the order of the activity instances for each participant, participant identification numbers, levels of reported pain, scores on the Hospital Anxiety and Depression Scale, and the self-report of MRSE as predictor variables. To account for uncertainty in imputation, we performed multiple single imputations, i.e. M single imputations with M > 1, with the decision tree. The hyperparameters of the classification tree and the feature set used were tuned for a primary imputed dataset. Hyperparameter setting was done with grid search, and the wrapper based selection technique employed in Study 2 was used to find the optimal feature set. These settings were also used for M = 5 secondary imputed datasets.

Similar to Study 2, leave-one-subject-out cross-validation was used to optimise the feature set and evaluate classification performance.

6.2 Results

In this section, we present the automatic detection performances, with each result based on aggregation of the automatic detection outputs from all folds.

Table 8 shows classification performance using decision trees, for the incomplete data (based on the use of the decision tree with surrogate splits) and the primary imputed dataset (without surrogate splits). Even without feature optimisation, imputation leads to better performance. Performance improves with feature set optimisation for the primary imputation and classification is well above chance level.

Classification performance using the SVM and the RF (without feature set optimisation) is given in Table 9. Their use does not lead to better classification performance than with the use of the decision trees (without feature set optimisation); in fact, the SVM and RF both perform less than the decision tree (with or without imputation) and their F1 scores for the lower level MRSE is worse than chance level classification.

With and without imputation and regardless of the classification algorithm used, there is better classification of the higher level MRSE class, likely because this is the dominant class.

Table 10 shows the results of classification for the secondary imputed datasets (using decision trees). For two of them, performance is well above chance level; in fact, one of these datasets has better performance than the performance with the primary imputed dataset.

We investigated if the models for the primary and secondary imputations performed worse for the Functional instances given that the data set was skewed towards Exercise instances. We tested this hypothesis by comparing classification accuracy for both subsets of instances across the six imputed datasets. A Wilcoxon Signed Rank Test indeed revealed a statistically significant difference in classification performance for the two movement types (z=-2.214, p<0.05, with large effect size r=-0.90). However, rather being lower, accuracy was significantly better for the Functional instances (median=0.80) than for the Exercise instances (median=0.71). This could be a result of lower intra-class variation for the Functional instances due to the fact that that subset of instances was smaller than the Exercise subset. However, we would argue that as reported in [61], low MRSE may have a stronger effect during functional movement, making the discrimination between low and high self-efficacy clearer in such settings than in controlled exercise settings

7 OVERALL DISCUSSION

Our investigation had two main objectives. First, it aimed to acquire an understanding of how physiotherapists, who are experts in human movement assessment, estimate MRSE through observation. The second objective followed from this and was to study how MRSE could then be automatically estimated from movement, with the long term aim of enabling technology with the capability to provide tailored support. We have described two studies that investigated this second objective in two physical activity contexts based on the cues that emerged from an initial study with physiotherapists. In this section, we discuss the implications of our findings and the opportunities that they present.

7.1 MRSE Features

An important contribution of our work has been to identify the relevant features that could contribute to building automatic MRSE level detection functionality. A major finding from Study 1 is that body movement is the main modality used by expert movement observers in MRSE assessment. While the importance of the body as a modality has been emphasised for affect in general [26][27] and for pain related affect in particular [28], only [18] had pointed to the relevance of body cues for MRSE specifically and then only in the context of choreographic performances. Our own finding points to the need for the tracking of body movement, to be able to automatically monitor MRSE levels. This has implications for the design of technology for physical rehabilitation, fitness, and sports, given the importance of MRSE to physical performance [7][11][12] and the growing use of technology in these areas [65]. For designers in these areas who may wish to incorporate MRSE detection functionality into technology, it will be necessary to integrate body movement sensors, particularly wearable body movement sensors that allow for monitoring in ubiquitous settings, in their design. The findings of our analysis of features in Study 2 additionally suggest that gross body movement such as measured by actigraphs or pedometers may not be sufficient. The increasing development of wearable sensors that track multiple anatomical segments such as Notch [66] point to the feasibility of body movement tracking for MRSE detection in these applications. Indeed, the findings from our investigation in Study 3 using a custom-built wearable sensing device support this hypothesis.

Another category of cues that was found in Study 1 to be used by expert movement observers to support MRSE assessment is engagement behaviour. Engagement behaviour was seen as engagement with people present during the exercise sessions and also engagement with the aspects of the surroundings that may affect the exercise (e.g. movement aids such as the chair). This cue could be particularly interesting when an avatar or robot is used as a physical rehabilitation coach (e.g. in [67]). Study 1 also suggests that facial expressions are informative for expert movement observers in assessing MRSE. However, a challenge in tracking facial expressions in the context of physical activity is the difficulty of capturing the face in such settings [68]. Nevertheless, in cases where physical activity is situated such as for certain physical exercises, facial expressions may be leveraged in improving automatic detection. Hence, an interesting challenge for the researchers working on facial expression recognition would be to investigate in depth what facial expressions could be related to MRSE. However, it should be noted that the literature on pain indicate that facial expressions are more common in the presence of others [28].

7.2 Automatic MRSE Level Detection

The main aim of our study was to investigate the possibility of automatically inferring MRSE levels in chronic pain physical rehabilitation. Findings in Study 2 provide ground-breaking evidence that self-efficacy for physical exercise movements can be automatically estimated from body movement as well as expert movement observers. As highlighted in our introductory section, this will enable technology provide rich and personalised support to people with chronic pain in the absence of a clinician. Such tailored intervention promises to promote adherence to useful strategies, promote confidence and independence, and foster engagement in physical activity [7][9][10][15][16]. While physical rehabilitation technologies are increasingly tracking various information from the user, physical performance is usually the main focus (e.g., in [69][70]). When psychological states are considered, mood and pain are the prime targets, such as in [70].

The findings from Study 3 show the feasibility of automatically detecting self-efficacy for functional movements in addition to physical exercises. This first step in investigating the feasibility of automatic MRSE monitoring in everyday physical functioning is important because of the necessity of providing tailored support in this context rather in physical exercises alone [9][16]. Even though physical exercise programs are designed to help people with chronic pain engage in everyday physical activities and the pursuit of important goals, physical exercises settings do not completely mirror everyday settings [10][16]. On one hand, physical exercises settings do not fully represent the complex demands of everyday functioning as they are typically done in controlled or artificial environments [10]. On the other hand, the value of a functional activity may compete with the psychological barrier associated with performing the movements that it involves or with adherence to helpful management strategies [16]. In addition, it has become evident in recent years that there is value in enabling incorporation of prescribed physical exercises into everyday physical functioning [9][16][71], e.g. performing stretches while washing dishes. This is because one of the reasons for non-adherence to prescribed exercises has been found to be poor availability outside of everyday routine [71]. Designs such as in [16][70][71] demonstrate the possibility of providing technological support in the context of everyday functioning, within and outside the home. One of the main challenges of automatic monitoring of movement related states in this context is the necessity of tracking body movement using portable, wearable body movement sensors. Findings in Study 3 provide evidence of the feasibility of automatic MRSE monitoring, outside the context of physical exercises, based on a minimal set of low-cost body movement sensors. A next step in extending this functionality would be the use of the prototype or similar to acquire body movement data during functional movements in the natural environment. Automatic MRSE monitoring in everyday functioning can then be properly validated.

7 Conclusion

For people with chronic pain, MRSE notwithstanding pain intensity influences engagement in physical activity and the pursuit of valued goals significantly [7]. Despite this knowledge and the discovery that MRSE feeds into clinical intervention [9], until now, there have been no attempts to investigate if and how technology may assess MRSE so as to be able to provide tailored intervention in the absence of a clinician. Our work addresses this gap. We provide an understanding of how clinicians estimate self-efficacy for physical exercise movements. We then show that technology can provide the same estimation by tracking movement behaviour alone, with F1 scores of 0.95 and 0.78 in two movement types. In addition, we provide evidence of the feasibility of estimating MRSE using a minimal network of low-cost body movement sensors and in functional movements as well as physical exercises, with F1 score of 0.79 in one movement type. Our work lays the groundwork that will enable physical rehabilitation technology for people with chronic pain directly address low MRSE. Our discussion highlights opportunities for more work in the area.

ACKNOWLEDGMENT

The project was funded by the EPSRC grant Emotion & Pain Project EP/H017178/1 and Olugbade was supported by the 2012 Nigerian Presidential Special Scholarship Scheme for Innovation and Development (PRESSID).

REFERENCES

[1] I. Tracey and M. Bushnell. (2009). How neuroimaging studies have challenged us to rethink. *J Pain. 10 (11)*, pp. 1113-20.

[2] H. Breivik, B. Collett, V. Ventafridda, R. Cohen, and D. Gallacher. (2006). Survey of chronic pain in Europe: Prevalence, impact on daily life, and treatment. *Eur J Pain. 10 (4)*, pp. 287-333.

[3] C. Johannes, T. Le, X. Zhou, et al.. (2010). The prevalence of chronic pain in United States adults. *J Pain. 11 (11)*, pp. 1230-9.

[4] M. Bushnell, M. Čeko, and L. Low. (2013). Cognitive and emotional control of pain and its disruption in chronic pain. *Nat Rev Neurosci. 14(7)*, pp. 502-511.

[5] J. Vlaeyen and S. Linton. (2000). Fear-avoidance and its consequences in chronic musculoskeletal pain. *Pain. 85(3)*, pp. 317-332.

[6] D. Turk and A. Okifuji. (2002). Psychological factors in chronic pain: evolution and revolution. *J Consult Clin Psychol. 70(3)*, pp. 678-690.

[7] A. Asghari and M. Nicholas. (2001). Pain self-efficacy beliefs and pain behaviour. A prospective study. *Pain. 94(1)*, pp. 85-100.

[8] A. Bandura. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review. 84(2)*, pp. 191-215.

[9] A. Singh, A. Klapper, J. Jia, A. Fidalgo, A. Tajadura Jimenez, N. Kanakam, N. Bianchi-Berthouze, and A. Williams, "Motivating people with chronic pain to do physical activity", in CHI, 2014, pp. 2803-2812.

[10] D. Turk, "Generalization and Maintenance of Performance," in Fordyce's Behavioural Methods for Chronic Pain and Illness, C. Main, F. Keefe, M. Jensen, J. Vlaeyan, K. Vowles, Ed. Washington, D. C., IASP Press, 2014, ch. 13, sec. Commentary, pp. 415-427.

[11] K. Gunter, J. Costa, K. White, K. Hooker, W. Hayes, C. Snow. (2003). Balance self-efficacy predicts risk factors for side falls and frequent falls in community-dwelling elderly. *J Aging Phys Act. 11(1)*, pp. 28-39.

[12] D. Treasure, J. Monson, and C. Lox. (1996). Relationship between self-efficacy, wrestling performance, and affect prior to competition. *Sport Psychol. 10(1)*, pp. 73-83.

[13] J. Martin and D. Gill. (1991). The relationships among competitive orientation, sport-confidence, self-efficacy, anxiety, and performance. *J Sport Exerc Psychol. 13(2)*, pp. 149-159.

[14] J. Cohn, T. Kruez, I. Matthews, Y. et al., "Detecting depression from facial actions and vocal prosody", in ACII, 2009.

[15] S. Felipe, A. Singh, C. Bradley, A. Williams, and N. Berthouze, "Roles for Personal Informatics in chronic pain", in PervasiveHealth, 2015.

[16] A. Singh, N. Berthouze-Bianchi, and A. Williams, "Supporting everyday function in chronic pain using a wearable device", in CHI, 2017.

[17] A. Belsi, E. Papi, A. McGregor. (2016). Impact of wearable technology on psychosocial factors of osteoarthritis management. *BMJ Open. 6*.

[18] J. Keogh, N. Griffin, and R. Spector. (1981). Observer Perceptions of Movement Confidence. *Res Q Exerc Sport. 52(4)*, pp. 465-473.

[19] D. V. Cicchetti. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment. 6(4)*, pp. 284-290.

[20] M. Aung, S. Kaltwang, B. Romera-Paredes, et al. (2016). The Automatic Detection of Chronic Pain-Related Expression. *IEEE T Affect Comput. 7(4)*, pp. 435-451.

[21] P. Watson, C. Booker, C. Main, and A. Chen. (1997). Surface electromyography in the identification of chronic low back pain patients. *Clin Biomech. 12(3)*, pp. 165-171.

[22] W. Janssen, H. Bussmann, and H. Stam. (2002). Determinants of the sit-to-stand movement: a review. *Phys Ther. 82(9)*, pp. 866-879.

[23] S. D'Mello and A. Graesser. (2009). Automatic detection of learner's affect from gross body language. *Appl Artif Intell. 23(2)*, pp. 123-150.

[24] K. McGraw and S. Wong. (1996). Forming inferences about some intraclass correlation coefficients. *Psychol Methods. 1(1)*, pp. 30-46.

[25] V. Braun and V. Clarke. (2006). Using thematic analysis in psychology. *Qual Res Psychol. 13(2)*, pp. 77-101.

[26] B. De Gelder. (2009). Why bodies?. *Phil. Trans. R. Soc. B. 364(1535)*, pp. 3475-3484.

[27] H. Aviezer, Y. Trope, and A. Todorov. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science. 338(6111)*, pp. 1225-1229.

[28] M. Sullivan, P. Thibault, A. Savard, R. Catchlove, J. Kozey, W. Stanish. (2006). The influence of communication goals and physical demands on different dimensions of pain behaviour. *Pain. 125(3)*, pp. 270-7.

[29] M. Karg, A. Samadani, R. Gorbet, et al. (2013). Body movements for affective expression. *IEEE T Affect Comput. 4(4)*, pp. 341-359.

[30] J. Wagner, F. Lingenfelser, E. Andre, and J. Kim. (2011). Exploring fusion methods for multimodal emotion recognition with missing data. *IEEE T Affect Comput. 2(4)*, pp. 206-218.

[31] K. A. Hallgren. (2012). Computing inter-rater reliability for observational data. *Tutor Quant Methods Psychol. 8(1)*, pp. 23-34.

[32] P. Werner, A. Al-Hamadi, K. Limbrecht-Ecklundt, et al. (2016). Automatic pain assessment with facial activity descriptors. *IEEE T Affect Comput.*

[33] N. Rathee and D. Ganotra. (2016). Multiview distance metric learning on facial feature descriptors for automatic pain intensity detection. *CVIU. 147*, pp. 77-86.

[34] J. Zhou, X. Hong, F. Su, and G. Zhao, "Recurrent convolutional neural network regression for continuous pain intensity estimation in video", in CVPR, 2016, pp. 84-92.

[35] J. Joshi, A. Dhall, R. Goecke, J. F. Cohn, "Relative body parts movement for automatic depression analysis", in ACII, 2013.

[36] H. Sarker, M. Tyburski, M. Rahman, et al., "Finding significant stress episodes in a discontinuous time series of rapidly varying mobile sensor data", CHI, 2016, pp. 4489-4501.

[37] Y. Matsuo, S. Miki, T. Takeda, et al., "Self-efficacy estimation for health promotion support with robot partner", ROMAN '15, pp. 758-762

[38] M. Aung, N. Bianchi-Berthouze, P. Watson, and A. Williams, "Automatic recognition of fear-avoidance behavior in chronic

pain physical rehabilitation", in PervasiveHealth, 2014, pp. 158-161.
[39] J. Dickey, M. Pierrynowski, DBednar, and S. Yang. (2002). Relationship between pain and vertebral motion in chronic low-back pain subjects. *Clin Biomech. 17(5),* pp. 345-352.
[40] T. Olugbade, M. Aung, N. Bianchi-Berthouze, N. Marquardt, and A. Williams, "Bi-modal detection of painful reaching for chronic pain rehabilitation systems", in ICMI, 2014, pp. 455-458.
[41] T. Olugbade, N. Bianchi-Berthouze, N. Marquardt, and A. Williams, "Pain level recognition using kinematics and muscle activity for physical rehabilitation in chronic pain", in ACII, 2015.
[42] P. Alborno, S. Piana, M. Mancini, et al., "Analysis of intrapersonal synchronization in full-body movements displaying different expressive qualities", AVI, 2016, pp. 136-143.
[43] S. Balasubramanian, A. Melendez-Calderon, and E. Burdet. (2012). A robust and sensitive metric for quantifying movement smoothness. *IEEE T Biomed Eng. 59(8),* pp. 2126-2136.
[44] D. Bernhardt and P. Robinson, "Detecting affect from non-stylised body motions", in ACII, 2007, pp. 59-70.
[45] S. Chen, Y. Tian, Q. Liu, and D. Metaxas. (2013). Recognizing expressions from face and body gesture by temporal normalized motion and appearance features. *Image and Vision Comput. 31(2),* pp. 175-185.
[46] H. Gunes, C. Shan, S. Chen, and Y. Tian, "Bodily expression for automatic affect recognition", in Emotion Recognition, 1st ed, Konar and Chakraborty, Eds., Wiley, 2015, ch. 14, pp. 343-377.
[47] J. Rivas, F. Orihuela-Espina, L. Sucar, L. Palafox, J. Hernández-Franco, and N. Bianchi-Berthouze, "Detecting affective states in virtual rehabilitation", in PervasiveHealth, 2015, pp. 287-292.
[48] M. Kächele, P. Werner, A. Al-Hamadi, et al., "Bio-visual fusion for person-independent recognition of pain intensity", MCS'15, pp. 220-230
[49] L. Breiman. (2001). Random Forests. *Mach Learn. 45(1),* pp. 5-32.
[50] C. Chen, A. Liaw, and L. Breiman. (2004). Using random forest to learn imbalanced data. UC Berkeley. California, USA.
[51] C. Cortes and V. Vapnik. (1995). Support-vector networks. *Mach Learn. 20(3),* pp. 273-297.
[52] K. Veropoulos , C. Campbell , and N. Cristianini, "Controlling the sensitivity of support vector machines", in IJAI, 1999.
[53] A. Janecek, W. Gansterer, M. Demel, et al., "On the relationship between feature selection and classification accuracy", in JMLR '08, pp. 90-105
[54] P. Narendra and K. Fukunaga. (1977). A branch and bound algorithm for feature subset selection. *IEEE Trans Comput. 26(9),* pp. 917-922.
[55] J. Jacobs, C. Yaguchi, C. Kaida, M. Irei, M Naka, S. Henry, and K. Fujiwara. (2011). Effects of experimentally induced low back pain on the sit-to-stand movement and electroencephalographic contingent negative variation. *Exp Brain Res. 215,* pp. 123-134.
[56] Invensense Inc. (2012). *MPU-9150 Product Specification Revision 4.0.* www.hobbytronics.co.uk/datasheets/PS-MPU-9150A.pdf
[57] Plux. (2015). *Electromyography (EMG) Sensor Data Sheet.* bitalino.com/datasheets/EMG_Sensor_Datasheet.pdf
[58] C. Duc, P. Salvia, A. Lubansu, et al. (2014). A wearable inertial system to assess the cervical spine mobility. *Med Eng & Phys. 36(1),* pp. 49-56.
[59] S. Simcox, S. Parker, G. Davis, R. Smith, and J. Middleton. (2005). Performance of orientation sensors for use with a functional electrical stimulation mobility system. *J Biomech. 38(5),* pp. 1185-1190.
[60] P. Roan, N. Deshpande, Y. Wang, B. Pitzer, "Manipulator state estimation with low cost accelerometers and gyroscopes", IROS, 2012.
[61] A. Singh, S. Piana, D. Pollarolo, G. Volpe, G. Varni, A. Tajadura-Jiménez, A. Williams, A. Camurri, and N. Berthouze. (2016). Go-with-the-flow. *HCI. 31(3-4),* pp. 335-383.
[62] A. Savitzky and M. Golay. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem. 36,* pp. 1627-1639.
[63] R. Little and D. Rubin, Statistical Analysis with Missing Data, 2nd ed, New Jersey: Wiley, 2002.
[64] L. Breiman, J. Friedman, C. Stone, and R. Olshen, Classification and Regression Trees, New York: CRC Press, 1984.
[65] R. Li, S. Kling, M. Salata, et al. (2016). Wearable performance devices in sports medicine. *Sports Health. 8(1),* pp. 74-8.
[66] Notch Interfaces Inc. Notch wearnotch.com/
[67] E. Wade, A. Parnandi, and M. Matarić, "Using socially assistive robotics to augment motor task performance in individuals post-stroke", IROS'15.
[68] R. Calvo and S. D'Mello. (2010). Affect detection. *IEEE T Affect Comput. 1(1),* pp. 18-37.
[69] C. Schönauer, T. Pintaric, H. Kaufmann et al., "Chronic pain rehabilitation with a serious game using multimodal input", in ICVR '11
[70] G. Duggan, E. Keogh, G. Mountain, et al. (2015). Qualitative evaluation of the SMART2 self-management system for people in chronic pain. *Disabil Rehabil Assist Technol. 10(1),* pp. 53-60.
[71] N. Bagalkot, S. Tomas, and B. Suraj. "ReRide", in PervasiveHealth, 2016.

Temitayo A. Olugbade is a researcher at University College London with interests in AI and cognitive processes and emotional states, particularly within the development of technology to assist, support, or augment human capabilities/experiences. Her PhD (in Affective Computing, at University College London) was funded by the Presidental Special Scholarship Scheme for Innovation and Development. In 2011, she was awarded a National Information Technology Development Fund for her MSc in Intelligent Systems (University of Sussex). She has been an IEEE Certified Biometrics Professional since 2010.

Nadia Bianchi-Berthouze is Professor in Affective Computing and Interaction. Her main area of expertise is the study of body posture/movement as a modality for recognising, modulating and measuring human affective states in human-computer interaction. She has published more than 170 papers in affective computing, human-computer interaction, and pattern recognition. She was awarded the 2003 Technical Prize from the Japanese Society of Kansei Engineering and she has been invited to give a TEDxStMartin talk (2012). She is/was: PI on the EPSRC-funded Emo&Pain project to design affective technology to support rehabilitation in chronic pain; co-I on the EPSRC-funded Digital Sensoria project investigating the use of biosensors to measure subjective responses to tactile experiences; co-I on the ILHAIRE project investigating laughter in human-avatar interaction; EU-FP7 Marie Curie IRSES UBI-HEALTH: Exchange of Excellence in Ubiquitous Computing Technologies to Address Healthcare Challenges, H2020 HUMAN Manufacturing, and HOLD funded by the Wellcome Trust.

Nicolai Marquardt is Senior Lecturer in Physical Computing at University College London. At UCL Interaction Centre he is working in the research areas of ubiquitous computing, physical user interfaces, proxemic interactions, and interactive surfaces. Nicolai received his PhD in Computer Science at the University of Calgary working with his advisor Saul Greenberg. He is a former intern at Microsoft Research Cambridge/UK and Redmond/USA and received an MSc/diplom degree in Media Systems from the Bauhaus-University Weimar. Nicolai is co-author of the book Sketching User Experiences: The Workbook (Morgan Kaufmann 2012) with Saul Greenberg, Sheelagh Carpendale, and Bill Buxton.

Amanda C. de C. Williams is Reader in Clinical Health Psychology at UCL, with a PhD from London University. She also works clinically in a pain management centre, and as research consultant for a human rights charity. She has worked in pain for about 30 years, particularly on evaluation of pain management interventions and their systematic review and meta-analysis. She has also developed research areas in evolutionary models for pain, behavioural expression of pain, and understanding of pain from torture.