Ultrasound-Based Detection of Fasciculations in Healthy and Diseased Muscles

Peter John Harding*, Ian D. Loram, Nicholas Combes, and Emma F. Hodson-Tole

Abstract—**Involuntary muscle activations are diagnostic indicators of neurodegenerative pathologies. Currently detected by invasive intramuscular electromyography, these muscle twitches are found to be visible in ultrasound images. We present an automated computational approach for the detection of muscle twitches, and apply this to two muscles in healthy and motor neuron disease-affected populations. The technique relies on motion tracking within ultrasound sequences, extracting local movement information from muscle. A statistical analysis is applied to classify the movement, either as noise or as more coherent movement indicative of a muscle twitch. The technique is compared to operator identified twitches, which are also assessed to ensure operator agreement. We find that, when two independent operators manually identified twitches, higher interoperator agreement (Cohen's κ) occurs when more twitches are present ($\kappa = 0.94$), compared to a lower number ($\kappa = 0.49$). Finally, we demonstrate, via analysis of receiver operating characteristics, that our computational technique detects muscle twitches across the entire dataset with a high degree of accuracy ($0.83 < \text{accuracy} < 0.96$).**

Index Terms—**Amyotrophic lateral sclerosis (ALS), computational analysis, diagnostics, image processing, ultrasound.**

I. INTRODUCTION

MOTOR neuron disease (MND) is a neurodegenerative disease which affects 2–3 per $100\,000$ people per year in the U.K. [1]. Diagnosis is based on the clinical observation of symptoms, including progressive weakness and wasting of muscles across several body regions [2]. Diagnostic certainty is expressed in terms of levels of probability and can be increased through the investigation of lower motor neuron function and indicators of chronic neurogenic change. Unstable or dying motor neurons fire intermittently, leading to involuntary activation of the group of muscle fibers they innervate, these events are defined as *fasciculations* [3]. Ongoing denervation results in the involuntary activation of individual fibers, these events are defined as *fibrillations* [4]. The activation of individual or groups of fibers results in localized twitches occurring in the muscle, which may be detected during a diagnostic examination. Both events can be found in healthy muscle, but greater numbers or distinctive patterns are diagnostic indicators of neurodegenerative disease [5].

A. Techniques for Detection of Involuntary Activation

In clinical practice, fasciculations and fibrillations are identified using electromyography (EMG), which detects action potentials (electrochemical changes in the muscle fiber membrane) resulting from activation. This typically involves inserting a needle electrode into the muscle of interest. For a diagnosis of the MND, fasciculation and fibrillation potentials must be detected across several body regions [6], so multiple needle insertions are required. The process is invasive and painful. In addition, the electrode detects activity in a small portion of muscle, meaning that many events in the muscle will not be detected. Alternatively, single and multichannel array electrodes can be placed on the skin to detect action potentials. This increases the volume of muscle which can be scanned, but it is limited to superficial muscle and is influenced by the distance between the muscle and the detecting electrode(s).

Ultrasound imaging (US) has been proposed for the detection of localized tissue movement, resulting from muscle activation during a fasciculation or fibrillation [4], [7]. US provides a noninvasive means of assessing muscle, enabling larger volumes of muscle, or multiple muscle layers to be assessed at once. US imaging has shown to be highly sensitive to movement, with movements as small as $5\,\mu m$ detectable [8], and most current machines can image at a temporal resolution of over 80 frames/s. US has previously been shown to be more sensitive than intramuscular EMG for detecting fasciculations in the tongue, biceps brachii (BB), and tibialis anterior, increasing the proportion of patients diagnosed with MND [7]. It was found that US identifies fasciculations in 80% of examinations ($71/89$) compared to a 45% detection rate with intramuscular EMG ($40/89$ examinations) [9]. A further study reports that US is more sensitive than EMG when applied to muscles which were clinically unaffected at the time of assessment [10]. Given the rapid physical deterioration which can occur with MND, early diagnosis is important for providing patients with appropriate care plans, and recruitment to clinical trials. The importance of detecting fasciculations in the early stages of the disease process has recently been highlighted [6], and interest in US image-based clinical assessment of muscles is growing, with calls for the development of prospective evaluation trials [11].

B. Steps Toward the Computational Analysis of Ultrasound Images

Studies of the application of US to the detection of involuntary muscle activation have only evaluated manual visual assessment

*P. J. Harding is with the Cognitive Motor Function Research Group, School of Healthcare Sciences, Manchester Metropolitan University, Manchester M15 6HB U.K. (e-mail: p.harding@mmu.ac.uk).

I. D. Loram and E. F. Hodson-Tole are with the Manchester Metropolitan University.

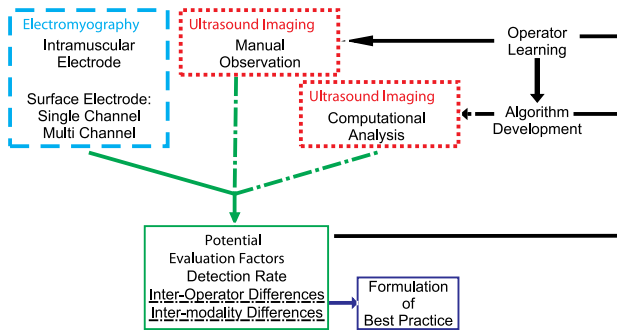N. Combes is with the Royal Preston Hospital.

Fig. 1. Schematic representation of sources of information bring combined to inform computational image analysis approaches. The assessment of visual techniques of detecting muscle twitches (Red, short dashed lines) are denoted by the pathway with the dashed line. In the future, myoelectric signals should be considered (Blue, large dashed lines). By combining different truth signals we are optimally placed to evaluate proposed computational analysis approaches, even where no single modality can reliably provide a "ground truth" signal.

of image sequences to determine the presence of fasciculations or fibrillations. These reports find acceptable intra- and interoperator reliabilities [4], [12], this approach limits the objective information gained from image sequences. Extracting more detailed information on the timing, duration, or frequency of the muscle twitches is time consuming and likely to introduce larger intra-/interoperator error and bias. Additionally, indicators of neurogenic change, e.g., fibrillations firing less than five times/s, which can be detected using EMG *are not visible* to human operators in US images [4], potentially contributing to reports that EMG is more sensitive than US for the detection of fibrillations [10]. It is clear that both EMG and US have disadvantages. No single modality currently provides details of all events which have occurred within a muscle or group of muscles during a given period of time.

There have been limited attempts to apply computational analysis approaches to the automated detection of muscle twitches in US, focusing on the identification of externally triggered muscle events [13]. This included the application of feature tracking across US sequences, with subsequent investigation of the sensitivity and specificity of mutual information (MI) and susceptibility metrics for the detection of electrically evoked muscle twitches in healthy individuals. Computational analysis will improve the value and application of this imaging modality as an objective diagnostic aid. Specifically, such approaches facilitate the extraction of quantitative information on muscle twitches, improving clinical diagnosis and monitoring, reducing the requirement for operator training, demands on operator time, and the potential for bias and errors. Evaluating the performance of any developed approach is complicated by the difficulties in defining a ground truth signal to compare results against. In such instances, development of a suitable computational approach needs to be evaluated through comparison with other available signals. This process is represented in Fig. 1, where signals from different EMG and manual observation of images can be compared and combined to develop best practice for sensitive, accurate, and timely identification of the occurrence of fasciculations and fibrillations. This process begins

by evaluating visual identifiers of fasciculations by comparing manual and computational methods of analyzing US images (the dashed pathway in Fig. 1). Specifically we provide: 1) evaluation of interoperator differences in manual identification of muscle twitches; 2) a method for computational analysis of US images to provide automated detection of muscle twitches; 3) comparison between computational and manual identifications.

II. Methods

A. Data Collection

US sequences were collected from two groups: 1) 20 healthy participants (10 female, 33 ± 13 years, 172 ± 8.6 cm, 73 ± 21.8 kg), recruited from the general population; and 2) Five participants who had been diagnosed with MND (two female, time since diagnosis: 3–18 months, 61.7 ± 15.7 years, 170.67 ± 6.03 cm, 75.33 ± 14.74 kg), recruited through the Motor Neuron Disease Care and Research Centre at the Royal Preston Hospital. All participants provided informed, written consent, and the study was approved by Ethics committees at Manchester Metropolitan University and Preston Royal Hospital.

In healthy participants, US sequences were collected (75 frames/s) from two muscles: the BB in the upper arm and the medial gastrocnemius (MG) in the lower leg. These muscles were chosen as they differ in the body region they are located, size and geometric characteristics (see Fig. 2) representing a range of properties found in muscles investigated during clinical examination. When studying BB, participants seated with their left arm placed in a slightly extended, supported, and relaxed position. The linear probe (7 MHz, 59 mm field of view, LogicScan 128, Telemed Ltd., Vilnius, Lithuania) was coated in acoustic gel and positioned to provide a clear longitudinal view. When studying MG, participants lay prone on a treatment couch, with their legs fully extended. The US probe was placed over the mid-muscle belly region of left MG. Participants remained as still as possible during collection of two 40-s trials. Where there was an obvious voluntary contraction, the data were deleted and the trial was repeated. The sequence in which muscles were studied was randomized in each participant. In MND-diagnosed participants, US images were also collected from BB and MG. Two 30–40 s sequences were recorded, using the same device outlined previously, with the probe held in place over the muscle of interest by an operator.

B. Operator Identification of Muscle Twitches

Operator identification of muscle twitches in collected US provides an *expert* truth signal, consisting of the frame number at which each twitch was identified as starting. Two operators, with over 11 years combined experience, independently screened all US. Each sequence was viewed on a frame-by-frame basis using video editing software (VirtualDub 1.10.4). Muscle twitches were defined as localized displacement of tissue and were, as reported in previous studies [9], easily distinguishable from other phenomena such as arterial pulsation (repeated, rhythmic displacements) and voluntary activation (movement of the whole muscle). Once a twitch had been identified, the operator closely
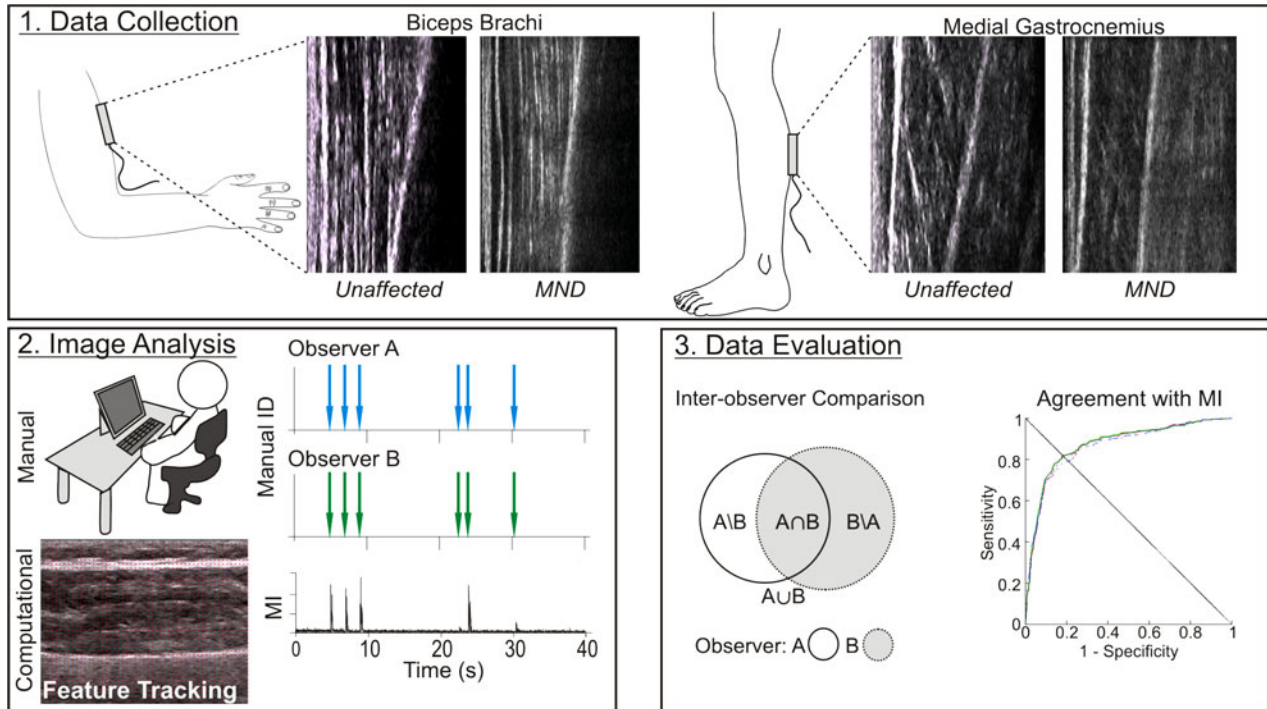
Fig. 2. Schematic representation of the flow of work. US was collected from each muscle across two groups of participants. US was assessed by two observers who noted the frame in which involuntary twitches began. US was also analyzed using MI to quantify characteristics of the movement patterns in tracked features. ROC are used to assess the accuracy of MI against operator observed events; Observer A only, Observer B only, $A \cup B$, and $A \cap B$.
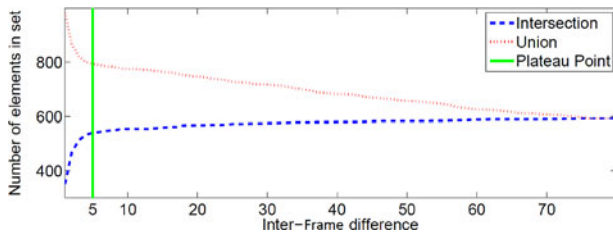

Fig. 3. Number of elements in $A \cap B$ and $A \cup B$ as the interframe difference is increased. The plateau point can be seen at approximately five frames.

screened frames around the time of initiation to identify the first frame in which tissue displacement was apparent. Operators did not report issues with voluntary activations being present in the data, although it is accepted that low-level sustained contraction may be difficult to identify (particularly in BB).

To assess the correctness and level of agreement between operators, the operator assessments were combined in several ways. Alongside the "raw" operator datasets, A and B, the intersection ($A \cap B$), and union ($A \cup B$) of the data were determined. To calculate the intersection of datasets, operators were classified as being "in agreement" over the presence of a muscle twitch when they had identified the twitch within five frames of each other. This figure was selected pragmatically, and represents the lowest interframe difference for which the sets $A \cap B$ and $A \cup B$ were insensitive to changes in interframe interval (see Fig. 3). Where discrepancies in frame number occurred between operators, within the five frame tolerance, the mean was used. Duplicated identifications in the union of data were removed using the same method.

This process resulted in a boolean vector, the same length as each video sequence, with true values representing each frame where a muscle twitch had been identified.

C. Computational Identification of Muscle Twitches

1) Tracking Muscle Tissue Displacements: The muscle tracking used here converts still images into vector grids approximating the movement between two consecutive images, the instantaneous movement identified at that image region between frames N and $N + 1$. The method employed for the detection of involuntary twitches are based on the Lucas–Kanade (LK) feature tracking algorithm [14], shown to accurately track movement within US images [13], [15]. Traditionally, a "first pass" of the image is made to identify features within the image that are mathematically superior for tracking, usually *corner* features are used, but previous work has shown that these are not effective for US images [15]. Therefore, an evenly spaced grid of features were used (grid size 100×80). The feature list is then stored as a set of *feature templates*, denoted as $T(x, y)$, where x and y represent the position of the templates' center pixel.

Using the same feature list throughout the analysis, an iterative search for each feature template from frame N is performed in frame $N + 1$ to find its position in the new image $I(x, y)$. At each iteration, the difference in intensity of $T(x, y)$ and $I(x, y)$ is calculated (ΔI), alongside the sums of the previous and current, horizontal, and vertical gradients which are used to estimate the error in both the x- and y-planes. The starting position for the search is updated according to these values, and Newton–Raphson iterations continue until the feature has net
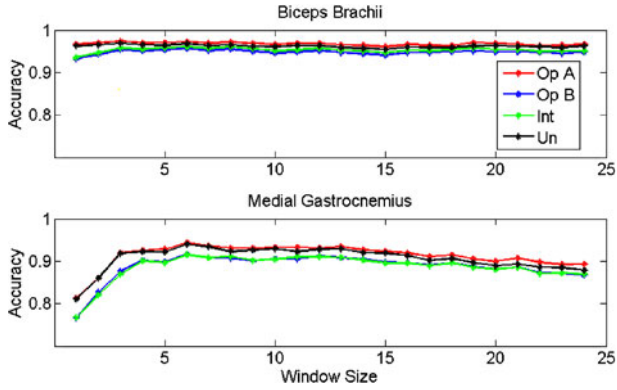
Fig. 4. Changes in accuracy values as the window factor (k) is increased, shown against all possible human generated truth signals; Op A and Op B (twitches identified by operator A or B), Int (twitches identified by both operator A and B), and Un (twitches identified by either operator A or B).
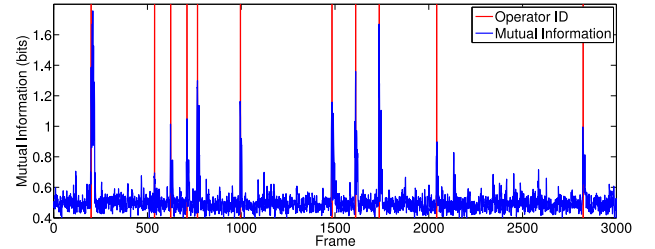

Fig. 5. Normalized output from an experiment on healthy subjects MG, showing the MI and observed fasciculations. Observed fasciculations were independently verified by two assessors. Parameters used were; truth signal ($A \cup B$), $\beta_{|m|} = 2$, $\beta_{\vec{m}} = \frac{\pi}{2}$, and $k = 5$.

movement of less than n pixels or ΔI becomes so great that the new region cannot be said to be the same feature. All features are replaced at every frame (using the original featurelist), ensuring an equal number of features for every tracking step.

Data obtained from each feature at every frame of the LK tracking are as follows; x and y (position of each feature, in pixels), and $\vec{m}$ (the feature's last movement vector).

The LK implementation used during this paper was a custom built, compute unified device architecture (CUDA) parallelized tracker. Parameters used during all analyses in this study were $F = 8000$ features, a feature window of 25×25, and minimum movement of $n = \frac{1}{10}$ pixel.

2) MI: The technique employed for the detection of fasciculations is MI, a statistical measure of the interdependence of two signals. This converts the 2-D grid of movement vectors, representing the movement between two frames, into a 1-D signal. The magnitude of the MI signal represents the likelihood of that US image transition containing a muscle twitch. We hypothesize that the magnitude of all feature movement ($|m|$) in images from quiet muscle would be small, mostly caused by the speckle noise present in all US images. As speckle noise is pseudorandomly distributed, the direction of movement $\vec{m}$ caused by this noise will also present randomly. Therefore, we suggest that under circumstances where there is no coherent movement, the interdependence of the magnitude and the direction of any feature movement would be unrelated and independently distributed. In contrast, during sequences where a muscle twitch is present, the magnitude of feature movement around that twitch would be greater, and the direction of movement more uniform and coherent. In short, the magnitude and the direction of movement within the local area surrounding a muscle twitch would be far more structured and more *predictable*.

The MI of $|m|$ and $\vec{m}$, denoted by $MI(|m|; \vec{m})$, yields a measure of the amount of information gained about the value of $\vec{m}$ if the value of $|m|$ is known and vice-versa. This means that, if all movements of specific magnitudes were all in the exact same direction then $MI(|m|; \vec{m}) \to \infty$, and if $|m|$ and $\vec{m}$ were entirely independent then $MI(|m|; \vec{m}) \to 0$. Therefore, we hypothesize that, by calculating the interdependence of these movement characteristics, it is possible to differentiate between

low-level noise and the involuntary twitch behavior we wish to identify.

MI is defined as:

$$MI(|m|; \vec{m}) = \sum_{i,j} P(|m|_i, \vec{m}_j) \log_2 \frac{P(|m|_i, \vec{m}_j)}{P(|m|_i)P(\vec{m}_j)} \quad (1)$$

where $P(|m|_i, \vec{m}_j)$, $P(|m|_i)$, and $P(\vec{m}_j)$ are the joint and individual probability distributions of $|m|$ and $\vec{m}$. As these probability distributions are discrete, there is some calibration required to find optimum discretization values for the continuous variables $\vec{m}$ and $|m|$, see Section III.

After application of the MI analysis, we are left with a 1-D signal, the magnitude of which represents the likelihood that a muscle twitch had occurred, see Fig. 5.

Each experiment run consisted between 2400 and 3000 frames of US, but the duration of a muscle twitch in US images ($\approx 0.5 \pm 0.11$ s) [16] is significantly greater than the temporal difference between frames ($0.012 < \tau < 0.013$ s). In the MI, muscle twitches were represented over multiple frames, rather than the single time points provided by the manual identification. Assuming that greater values of MI are indicative of muscle twitches, any potential twitch must present as a "peak." Leading from this, the "start" and "end" of the twitch were respectively identified as the first point before and the first point after the peak which lay below the noise threshold. The noise threshold was taken as the signal mean/trial.

The result of the entire MI analysis process is a discrete set of possible muscle twitch events, each having a start frame, end frame, and magnitude.

D. Evaluation

To assess agreement between two independent observers, the κ statistic was employed. A κ statistic is a means of assessing the level to which two or more classifiers (human, computational, etc.) agree upon the classification of certain data. The meaning of the κ value is described in terms of five ranges: Slight $0.0 \leq \kappa < 0.2$, fair $0.2 \leq \kappa < 0.4$, moderate $0.4 \leq \kappa < 0.6$, substantial $0.6 \leq \kappa < 0.8$, almost perfect $0.8 \leq \kappa \leq 1.0$ [17].

Receiver operating characteristics (ROC), commonly used to assess the accuracy of binary classifier systems, were used to compare the MI signal to the human generated truth signal. In the present analysis, the human identified twitches were used to create four different truth signals: 1) twitches identified by

operator A, 2) twitches identified by operator B, 3) twitches identified by both operators ($A \cup B$), and 4) twitches identified by either operator ($A \cap B$).

To calculate the ROC, peaks whose boundaries encompassed a manual identification of a twitch were counted as *true*. If there was no corresponding manual identification within the duration of the peak, the peak was counted as *false*. See Fig. 5, for example, MI and operator IDs. The *accuracy*, as reported for the entirety of this paper, is defined as the area under an ROC curve.

III. CALIBRATION

To enable the parametrization of the image analysis approaches, data from a subset of five, arbitrarily chosen, healthy participants were selected to provide a calibration dataset (two female). *These data were used for the calibration of the MI and the window factor only and do not form part of the test data nor are they included in the final results.* Only data from healthy participants were used for this purpose, to retain the maximum size of the clinical population dataset.

A. Window Factor

US images are known to exhibit noise, and during a motion tracking task the signal-to-noise ratio (SNR) may be very small. The SNR may be increased by dropping the temporal resolution of the US stream (effectively changing the frame rate), this amplifies the movement between two frames, but has little effect on noise. We term the distance between tracked frames as the *window factor*, and calibration tests were carried out to investigate whether increases in the window factor could be used to improve the SNR. Rather than tracking features from frame $N \rightarrow N+1$, multiple tracking targets are used. Features selected in frame N are stored, and were tracked into k subsequent frames. These metrics are analyzed to investigate the effects of window factor on ROC accuracy across both muscles and population groups. Values of k were tested up to a total of $k = 25$, approximately 0.30–0.33 se of real time, and the accuracy of classification against each truth signal were recorded, see Fig. 4.

The window size k was shown to have a large effect on data recorded from MG, with values of ≈ 0.8–0.9 for $k \leq 3$ before rising to 0.9 at $k \geq 5$. In comparison, k has little effect on the BB data. For the purposes of the work presented here $k = 5$ *is used for all further analysis.*

B. Discretization of MI Parameters

Discretization of the speed $|m|$ and direction $\vec{m}$ of feature movement was tested across multiple values to find an optimum point at which the difference between quiet and twitch behavior could be distinguished. As the constraint on feature movement per frame was limited to 16 pixels, the following bin sizes were assessed: $\beta_{|m|} \in \{8, 4, 2, 1\}$, and the same for $\beta_{\vec{m}} \in \{\frac{\pi}{8}, \frac{\pi}{4}, \frac{\pi}{2}, \pi\}$.

Results from this analysis of images collected from BB and MG are shown in Table I. For the assessment of BB data $\beta_{|m|} = 8$

TABLE I
CHANGES IN MI DISCRETIZATION PARAMETERS AND THE RESULTING ROC ACCURACY, WHEN COMPARED TO OPERATOR IDENTIFICATION IN BB (TOP) AND MG (BOTTOM)

		$\beta_{|m|}$ pixels			
		8	4	2	1
$\beta_{\vec{m}}$ c	$\frac{\pi}{8}$	0.7566	0.8941	0.8770	0.8039
	$\frac{\pi}{4}$	0.8445	0.9095	0.9095	0.8445
	$\frac{\pi}{2}$	0.8039	0.8770	0.8770	0.8770
	π	0.9629	0.8941	0.8770	0.8770
		$\beta_{|m|}$ pixels			
		8	4	2	1
$\beta_{\vec{m}}$ c	$\frac{\pi}{8}$	0.6162	0.6927	0.6749	0.6779
	$\frac{\pi}{4}$	0.6765	0.6972	0.7129	0.6990
	$\frac{\pi}{2}$	0.6321	0.7295	0.8108	0.8050
	π	0.6605	0.7301	0.7244	0.7653

and $\beta_{\vec{m}} = \pi$ provided the greatest accuracy, while in MG $\beta_{|m|} = 2$ and $\beta_{\vec{m}} = \frac{\pi}{2}$ was optimum.

These calibration values are physiologically meaningful, as they can be directly related to the structure of the two different muscles. The fascicles in the BB are arranged in parallel, i.e., they run longitudinally along the muscle, meaning the contraction and relaxation of fibers would occur horizontally across US images. This change is unlikely to be viewed directly, but the overall expansion of the muscle, which would primarily present vertically would. Evidence of this can be seen in the optimal calibration parameter $\beta_{\vec{m}} = \pi$, which in effect classifies all movement into either *toward* or *away* from the probe, disregarding horizontal movement in the image. The fascicles in the MG however are *pennate*, i.e., attached obliquely, so they appear diagonally in the US image, this means that contractile movement presents both vertically and horizontally. Again, evidence of this can be seen in the calibration parameter $\beta_{\vec{m}} = \frac{\pi}{2}$, which would take into account both horizontal and vertical movements in the US images.

The values stated here are the discretization values used for *all* further calculations presented in this paper.

IV. RESULTS

A. Interoperator Agreement

The interoperator agreement across the dataset was found to be generally high, with a distinct positive trend between the number of twitches present per experiment and the resulting agreement between operators. Statistics on interoperator agreement can be seen in Table II.

We see that the muscles and population groups where twitches were most prevalent show the highest agreement between operators; this is in contrast to the low agreement in the BB of the healthy population, where twitches were rare. Differences in

TABLE II
STATISTICS ON OPERATOR IDENTIFICATION OF MUSCLE TWITCHES, $(A \cap B)$
WAS CALCULATED USING AN INTER-FRAME DIFFERENCE OF FIVE FRAMES

	Healthy		MND	
	BB	MG	BB	MG
A	36	116	169	156
B	40	151	183	151
$A \cap B$	17	102	167	139
κ	0.49	0.76	0.94	0.85

the number of twitches identified by each operator were found, with the more experienced operator A classifying more twitches overall, but with just two operators there is insufficient data to draw significant comparisons.

The occurrence of muscle twitches was found to be greatly different across both the healthy and MND populations, and the BB and MG muscles, which can be seen by analyzing the union $(A \cup B)$ of the two operator IDs. The average occurrence of muscle twitches in the healthy dataset was found to be: MG ($\bar{x} = 0.12\,\text{s}^{-1}$, $\sigma = 0.15\,\text{s}^{-1}$) and BB ($\bar{x} = 0.03\,\text{s}^{-1}$, $\sigma = 0.04\,\text{s}^{-1}$). The average occurrence of muscle twitches in the MND dataset was found to be: MG ($\bar{x} = 0.54\,\text{s}^{-1}$, $\sigma = 0.30\,\text{s}^{-1}$) and BB ($\bar{x} = 0.56\,\text{s}^{-1}$, $\sigma = 0.52\,\text{s}^{-1}$). We see from this that the mean occurrence of muscle twitches in the pathological MG data was nearly five times greater than in the healthy muscle, while the BB showed that pathological muscle exhibited greater than 18 times more muscle twitches than the healthy cohort. We believe this to be because the MG muscles are far more likely to exhibit nonpathological fasciculations, leading to the difference between healthy and MND affected muscle to be less pronounced.

B. Healthy Population

The agreement between observer identification of twitches in healthy participants differed between the two muscles studied (see Table II). The MG showed the best agreement ($\kappa = 0.76$), categorized as substantial agreement. Agreement between operators for BB was lower ($\kappa = 0.49$), categorized as moderate. The number of muscle twitches in the MG and BB differed greatly, with 76 twitches identified in the BB, whereas 267 were identified in the MG (see Table II).

When evaluating the agreement between observer identified muscle twitches and MI, the accuracy again differed between muscles, but also between truth signals (see Figs. 2 and 6). The differences between the two observers when analyzing the healthy data were more pronounced. Comparison against A had the better accuracy in each muscle (BB 0.94; MG 0.91), while B showed lower agreement (BB 0.93; MG 0.89).

C. MND Affected Participants

Both observers identified greater numbers of involuntary muscle twitches in image sequences recorded from BB and MG in MND affected participants (see Table II, note $N = 5$ versus $N = 15$ for affected and healthy participants, respectively). The agreement was classified as almost perfect ($\kappa > 0.8$), although
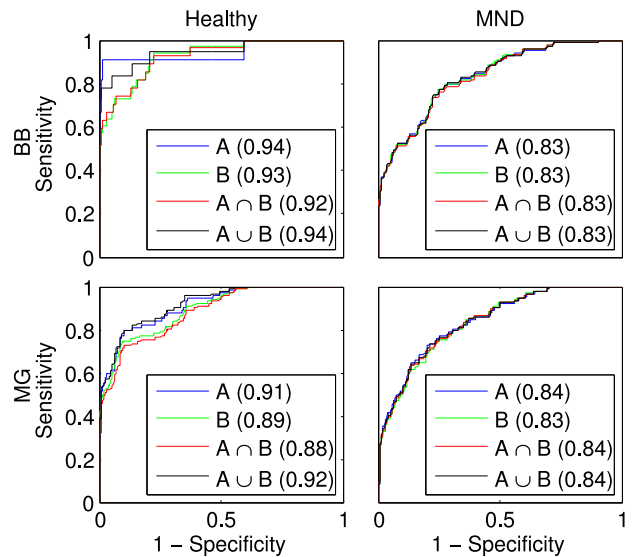

Fig. 6. ROC curves for all muscles, across both populations, showing single and combined operator identification performances (A, B, $A \cup B$, and $A \cap B$). Accuracy values presented in figure legends.
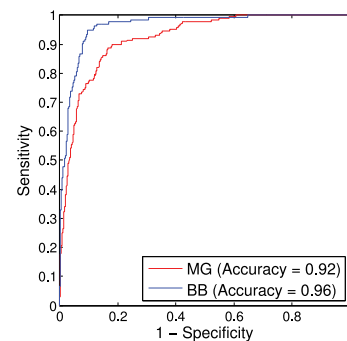

Fig. 7. ROC curves produced when healthy and affected groups were combined.

differences did occur between the two muscles with the greatest agreement in BB ($\kappa = 0.94$) compared to ($\kappa = 0.85$) in MG.

When evaluating the agreement between observer identified muscle twitches and the MI metric, few differences were observed between the muscles studied and the truth signal (see Fig. 2). In all cases the highest accuracy occurred in the BB muscle. Using individually identified events ($A \cup B$) provided a marginally higher agreement than either set of individual observations.

D. Combined Data

Comparison of MI and the intersection ($A \cap B$) in both healthy and affected participants is shown in Fig. 7. Accuracy in both muscles was high, slightly better in BB (0.96 versus 0.92).

V. DISCUSSION

This paper has met three key objectives: 1) evaluation of operator agreement in the identification of muscle twitches, 2) a computational method for the detection of muscles twitches, and 3) comparison between manual and computational twitch

identification. Across the whole dataset, κ statistics revealed almost perfect agreement between the two operators (see Table II). The highest level of agreement occurred in the clinical population, where the largest number of twitches were present. The poorest agreement was in the BB of healthy participants, where the lowest number of twitches were identified. Agreement appears to be strongly influenced by the number of twitches present in the data being analyzed, fewer twitches corresponding to lower agreement in all cases.

Previous work evaluating interoperator agreement reported a figure of $\kappa = 0.84$ for two operators viewing data from 317 muscles across 13 patients [12]. It did not provide insight into the variation of agreement across muscles studied or how agreement between operators was defined (i.e., on the basis of individual twitch events or presence/absence of twitches). The only other work which evaluates operator agreement focuses on the identification of fibrillations in collected images [4]. Therefore, agreement was influenced by operator experience, with more experienced operators having greater agreement ($\kappa \approx 0.63$–0.67) than the least experienced ($\kappa \approx 0.22$–0.55). This study only determined the presence or absence of fibrillations, ignoring more objective measures (e.g., number of events, rate of occurrence, morphology of twitches) which could provide valuable diagnostic and monitoring information [6]. Future work will investigate whether this information may be automatically extracted from US images. Such measures have previously been time consuming, with analysis of each video from healthy and affected participants presented here taking approx. 300–400 and 660–850 s, respectively. Our results, coupled with those of others [4], suggest that a detailed assessment of both the training requirements for, and sensitivity of, manual screening is imperative if US is to progress as a diagnostic tool for neurodegenerative disease.

We presented the first automated approach for the detection of involuntary muscle twitches via US. Comparison of MI with manual identification of twitches revealed a high level of classification accuracy (see Fig. 6). The results confirm the potential of computational analysis as a means of objectively determining the occurrence of fasciculations. MI was selected for testing based on our previous study with electrically evoked muscle twitches [13] and we have shown it to be an effective detector of the smaller, involuntary activations; part of current diagnostic criteria [6].

The agreement found between manual and computational validates the MI approach as a diagnostic tool, but further study is required to determine the wider application of this approach. Consideration of more muscles, as clinical tests screen muscles across multiple body regions (e.g., upper/lower limbs, trunk, head, and neck). Finally, truth signals are required for the evaluation of algorithms. Future work will include collection of EMG to evaluate accuracy. This protocol will enable investigation of the potential to automatically detect/classify other involuntary activation events (e.g., fibrillations, or complex repetitive discharges) and enable determination of the optimal combination of assessment approaches for diagnosis. The potential to characterize other features of muscle twitches (e.g., spatial and temporal features) should not be overlooked, as it will provide additional information for diagnosis and monitoring of disease progression, of particular value for the development of new therapeutic interventions.

VI. Conclusion

We have shown that the computational analysis of US images provides a means of automatically detecting involuntary activations in muscles of both healthy and MND affected individuals. MI-based analysis of tracked LK image features shows strong potential as an approach to provide automated objective measures from US images. Further work is however required to determine its efficacy across a wider range of muscles and such work should also consider potential application of other analysis approaches.

References

[1] S. Bourke *et al.*, "Non-invasive ventilation in MND: Current UK practice," *ALS MND: Off. Publication World Federation Neurology, Res. Group MND*, vol. 3, no. 3, pp. 145–149, 2002.

[2] M. de Carvalho *et al.*, "Electrodiagnostic criteria for diagnosis of ALS," *Clin. Neurophys.*, vol. 119, no. 3, pp. 497–503, 2008.

[3] S. Pillen *et al.*, "Muscle ultrasound in neuromuscular disorders," vol. 37, pp. 679–693, 2008.

[4] S. Pillen *et al.*, "Muscles alive: Ultrasound detects fibrillations," *Clin. Neurophysiol.*, vol. 120, no. 5, pp. 932–936, 2009.

[5] J. a. Costa *et al.*, (Nov. 2012). Awaji criteria for the diagnosis of amyotrophic lateral sclerosis:a systematic review. *Archives Neurol. 69(11)*, pp. 1410–6. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/22892641

[6] M. de Carvalho and M. Swash. (2013). Fasciculation potentials and earliest changes in motor unit physiology in ALS. *J. Neurol., Neurosurg., Psychiatry* [Online]. *84(9)*, pp. 963–968. Available: http://www.ncbi.nlm.nih.gov/pubmed/23418210

[7] S. Misawa *et al.*, "Ultrasonographic detection of fasciculations markedly increases diagnostic sensitivity of ALS," *J. Neurol.*, vol. 77, no. 16, pp. 1532–1537, 2011.

[8] I. D. Loram *et al.*, "Use of ultrasound to make non-invasive in vivo measurement of continuous changes in human muscle contractile length," *J. App. Physiol.*, vol. 100, no. 4, pp. 1311–1323, 2006.

[9] I. M. P. Arts *et al.*, "Muscle ultrasonography: A diagnostic tool for ALS," *Clin. Neurophys.*, vol. 123, no. 8, pp. 1662–1667, 2012.

[10] A. Grimm *et al.*, "Muscle ultrasonography as an additional diagnostic tool for the diagnosis of ALS," *Clin. Neurophys.*, vol. 126, no. 4, pp. 820–827, 2015.

[11] M. Swash and M. de Carvalho, "Muscle ultrasound detects fasciculations and facilitates diagnosis in ALS," *Neurology*, vol. 77, no. 16, pp. 1508–1509, 2011.

[12] C. D. Reimers *et al.*, "Fasciculations: Clinical, electromyographic, and ultrasonographic assessment," *J. Neurol.*, vol. 243, no. 8, pp. 579–584, 1996.

[13] P. J. Harding *et al.*, "Automated detection of skeletal muscle twitches from b-mode ultrasound images," in *Proc. IEEE 21st Int. Conf. Pattern Recognit.*, 2012, pp. 2630–2633.

[14] B. D. Lucas *et al.*, "An iterative image registration technique with an application to stereo vision," in *Proc. Int. Joint Conf. Artif. Intell.*, 1981, vol. 81, pp. 674–679.

[15] R. Cunningham *et al.*, "Automated measurement of human skeletal calf muscle contraction via b-mode ultrasound imaging," presented at the 17th Int. Conf. Medical Imaging, Understanding Analysis, Birmingham, U.K., 2013.

[16] F. O. Walker *et al.*, "Sonographic imaging of muscle contraction and fasciculations: A correlation with electromyography," *Muscle Nerve*, vol. 13, no. 1, pp. 33–39, 1990.

[17] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977.