UNIVERSITY OF BATH

University of Bath

TITLE PAGE

A novel automatic method for monitoring Tourette motor tics through a wearable device.

[1,2]Michel Bernabei, [1,2,✉]Ezio Preatoni, [1,3]Martin Mendez, [2]Luca Piccini, [4]Mauro Porta, [2]Giuseppe Andreoni

[1]*Dipartimento di Bioingegneria, Politecnico di Milano, Milan, Italy*
[2]*Dipartimento di Industrial Design, Arti, Comunicazione e Moda (INDACO), Politecnico di Milano, Milan, Italy*
[3]*Facultad de Ciencias, Universidad Autónoma de San Luis Potosí, San Luis Potosí, Mexico*
[4]*Department of Neurology, Tourette Centre, IRCCS "Galeazzi", Milan, Italy*

✉**Corresponding author:**
Ezio Preatoni, PhD
Dipartimento INDACO, Politecnico di Milano
Piazza Leonardo Da Vinci, 32 - 20133 Milano, Italy
Phone: +39 (02) 2399-3357/5999
ezio.preatoni@polimi.it

Word count: 2633

Running Title: A novel method for monitoring Tourette motor tics

ABSTRACT

The aim of this study was to propose a novel automatic method for quantifying motor-tics caused by the Tourette Syndrome (TS).

In this preliminary report, the feasibility of the monitoring process was tested over a series of standard clinical trials in a population of 12 subjects affected by TS. A wearable instrument, with an embedded three-axial accelerometer was used to detect and classify motor tics during standing and walking activities. An algorithm was devised to analyze acceleration data by: eliminating noise; detecting peaks connected to pathological events; classifying intensity and frequency of motor tics into quantitative scores. These indexes were compared with the video-based ones, provided by expert clinicians, which were taken as the gold-standard. Sensitivity, specificity and accuracy of tic detection were estimated, and an agreement analysis was performed through the least square regression and the Bland-Altman test.

The tic recognition algorithm showed sensitivity=80.8±8.5% (mean±sd), specificity=75.8±17.3%, accuracy=80.5±12.2%. The agreement study showed that automatic detection tended to overestimate the number of tics occurred. Though, it appeared this may be a systematic error due to the different recognition principles of the wearable and video-based systems. Furthermore there was substantial concurrency with the gold-standard in estimating the severity indexes.

The proposed methodology gave promising performances in terms of automatic motor-tics detection and classification in a standard clinical context. The system may provide physicians with a quantitative aid for TS assessment. Further developments will focus on the extension of its application to everyday long term monitoring out of clinical environments.

Key Words: Tourette syndrome; automatic tic detection, accelerometers, wearable monitoring.

INTRODUCTION

Tourette Syndrome (TS) is a chronic neurologic disorder characterized by the childhood onset of multiple motor and phonic tics that wax and wane over time [1,2]. A distinctive issue concerning the assessment of TS is the difficulty in quantifying and classifying objectively its various clinical manifestations. In fact, the wide range of visible and audible signs may induce subjectivity in the evaluation process, in spite of the rules set by literature for defining TS scores. Moreover: (i) symptoms vary unpredictably over time [3]; (ii) patients are able to suppress in part or totally their symptoms for minutes to hours [3]; (iii) situational stimuli can change tic expression [3,4,5]. Multiple variables such as frequency, number of tic-types, intensity, complexity, body distribution, suppressibility, and interference with normal activities are commonly considered to assess the severity of TS.

The evaluation of TS is currently carried out either through clinical examination or through patient reports based on self-assessment of tic disorder. Both methods have strengths and weaknesses: patients' reports are highly subjective and may be substantially different from the physician's evaluation; clinical scores are a semi-quantitative measure, but still depend on the examiner opinion and experience.

Goetz et. al. introduced a new method to evaluate TS symptom severity through a video-based tic rating scale [6]: specific domains of tic characterization (severity, frequency, body distribution) are examined through visual observation of video-recordings and assessed through the attribution of diagnostic scores correlated with well-known clinical severity scales (Yale Global Tic Severity Scale, YGTSS) [7]. The analysis of video-recordings is the standard reference for many TS studies, though it also lacks in objectivity and may lead to different medical interpretations. Therefore an automatic system based on wearable technologies and methods for the quantitative assessment of motor disorders may help to overcome this problem.

Wearable technologies (WT) have been exploited for gathering biological data in the long-term monitoring field [8-15]. They have been used to recognize motor activities [10,11], to define pathological features in Parkinson's disease [12,13], and to observe patients during rehabilitation [14,15], giving the opportunity to reach mid or long-term data recordings in both clinical and home environments.

Accelerometers are currently among the most widely studied wearable sensors for activity recognition, thanks to their accuracy in the detection of human body movements, small size, and reasonable power consumption [16]. Accelerometers may allow for continuous observation periods, suited to the changeable nature of the TS, without affecting the patient's daily activities. Despite their potentialities, methods based on wearable accelerometers have still not been exploited for the evaluation of TS motor disorders, probably because of the difficulty in recognizing tics during the execution of other motor activities.

Therefore, the aim of this study was to perform a first preliminary study on a novel wearable monitoring system for the automatic assessment of TS. The feasibility study was carried out on a sample of TS subjects to understand potentialities and applicability in the clinical environment. At this stage, we focused onto the stronger motor manifestations of TS, leaving to future refinements the issue of vocal tics, slight facial events or dystonic signs. We estimated the reliability of the automatic scoring compared with the video-based one (gold-standard).

METHODS

Subjects

Twelve TS patients (10 males and 2 females, age between 17-45 years) participated in the current study. The population presented a heterogeneous typology of motor tics, variable in body distribution, intensity, frequency and complexity. All patients were pharmacologically treated with neuroleptics or botulinum toxin and four of them had a Deep Brain Stimulation (DBS) implant.

The study was approved by the competent Institutional Review Board. Subjects were properly informed about testing procedures, personal data treating and aims of the research, and they provided informed consent before participation.

Instrumentation

An actigraph based on a commercial system was used to detect pathological events (PROTHEO I, SXT − Sistemi per Telemedicina). The device consists in a plastic case (92x58x25 mm) containing a 3D acceleration sensor (LIS3L06AL, STMicroelectronics), a bluetooth® transmission module (PAN1540, Panasonic) and a rechargeable LiIon battery.

The wearable device was supported by a software designed for collecting and manage data in applications requiring real-time monitoring of biosignals and movement (HIM, Sensibilab, Politecnico di Milano).

Acquisition Protocol

The videotape protocol proposed by Goetz et al. [17,18] involves 4 sessions of about 2.5 minutes. The patient is placed in a quiet room in front of a video-camera. Two body views are recorded, full frontal body, and head and shoulders only under two conditions, relaxed with the presence of the examiner, and relaxed with the patient alone. For this study we focused on the comparison between

the automatic system and the video-based procedure, referring to the full frontal body observation in presence of the examiner. Two characteristics of the motor-tics were evaluated: intensity and frequency.

The main challenge of the process was to discriminate tics from normal movements only through accelerations. We asked subjects to perform two motor tasks: standing, to evaluate the system performance when the tic acceleration is remarkably identifiable from the base signal; walking, to assess it when tic acceleration may be hidden by normal movements waveforms. Six-minutes acquisitions were performed for each subject (2 trials of 3 minutes each). The device was positioned with a band on the chest for the early 3 minutes and on the dorsal area (L2-L3 level) for the last 3 minutes. The trial session was defined as:

- 1 min standing, with arms close to the body sides;

- 1 min random walking across the room;

- another 1 min standing stage.

Motion data were recorded by a computer near the trial location. Two kinds of data were collected: videotape recordings and triaxial acceleration signals. Acceleration data were calibrated, digitized and recorded by a remote processing unit through bluetooth® transmission (fig. 1-A).

Data Processing

The videotape recordings were submitted to an experienced physician that rated a tic severity score focusing on intensity and frequency domains of the video-based scales proposed by Goetz et al. [6]. The rater was blinded for acceleration data. Two severity scores were provided for each subject: I_{ST}, index of intensity, and F_{ST}, index of frequency.

To our knowledge, there is no previous information in literature regarding the analysis of acceleration signals for the conditions shown by TS. In order to detect the intensity and frequency of the tic, processing of acceleration signals was designed to discriminate symptomatic episodes from common activities, which in this application represent an unwished "noise".

6

The algorithm was specifically developed to analyse TS activity, was implemented in Matlab® (Mathworks Inc.), and consisted of 2 main steps, which were both independent from any video reference and thus made the procedure completely automatic.

The first step, "recognition", aimed to maximize tic acceleration spikes in comparison with normal-movements accelerations. Spikes are fast changes in the acceleration signal and correspond to abrupt and instantaneous changes in postures or natural movements (e.g. walking). We subsequently carried out: a non-linear median filtering; a NEO filtering (Non-linear Energy Operator [19]); and an adaptive thresholding. The median filter was used to identify and subtract the baseline from the acceleration signal. The NEO filter was used to emphasize acceleration spikes over the waveform baseline that represents a noisy background. The adaptive thresholding was used to discriminate spikes due to tic events from the ones determined by normal movements. It was based on the variance of the spectral distribution of the subject's motion activity (fig. 1-B).

The second step, "classification", assigned intensity and frequency scores (I_{NEW}, F_{NEW}) to the tic activity of each subject according to preset ranges of, respectively, the absolute measure of the acceleration, and the number of events over time (fig. 1-C). I_{NEW} and F_{NEW} ranged between 0 and 4 and were implemented on the base of the well-known scores expressed in the video-based tic rating scale proposed by Goetz et al. [6] and in the YGTSS [7].

The intensity scale, in g=9.81 m/s:

0. barely perceptible: $0 \leq I_{new} < 0.5$;

1. visible: $0.5 \leq I_{new} < 1$;

2. some problem: $1 \leq I_{new} < 1.5$;

3. impaired function: $1.5 \leq I_{new} < 2$;

4. no function: $I_{new} \geq 2$.

The frequency scale, in tics/s over 2s-windows:

0. no tics;

1. $0 \leq F_{new} < 0.083$;

2. $0.083 \leq F_{new} < 0.16$;

3. $0.16 \leq F_{new} < 0.25$;

4. $F_{new} \geq 0.25$.

The I_{NEW} scale intervals were designed to compensate the sensitivity gap between the accelerometric device and a human scorer. I_{NEW} ranges were set referring to the highest and the lowest severity case occurred inside the trial population.

**** Insert Figure 1 near here ****

The algorithm parameters for automatically detecting motor-tics were chosen to maximize sensitivity (SEN), specificity (SPE) and accuracy (ACC) in respect of the tics recognized by the human expert. Videotape analysis allowed to identify the timeline of actually occurred tics (fig. 1). Differences in performance between the frontal and dorsal positioning of the actigraph were assessed by applying Mann-Witney tests (P=0.05) to SEN, SPE and ACC (tab. 1).

We compared scores from the automatic system (I_{NEW}, F_{NEW}) and from the gold-standard counterparts (I_{ST}, F_{ST}) by counting the number of concordances and by using Wilcoxon tests (P=0.05). This comparison was useful to assess the agreement between the quantitative measure of the device and the qualitative rating system of the physician's analysis.

Furthermore we performed a least-square regression analysis and a Bland-Altman (B-A) plot [20] on frequency values (in tics/s) from our methodology and from the gold-standard. The correlation and the B-A analyses were not applied to the intensity index because the video-based system does not provide quantitative measures in terms of accelerations, but only a qualitative observation and a 0-4 ordinal classification.

8

RESULTS

Sensitivity, specificity, accuracy of the tic recognition algorithm showed mean values that ranged between 75.8 and 80.8% for both the adopted configurations (tab. 1). Results showed that the evaluated positions did not interfere with tic recognition ability: percentage differences between frontal and back positioning were about 1.9, 2.5, 0.1% for mean SEN, SPE and ACC; Mann-Witney tests did not evidence significant changes between the two conditions (P=0.77, P=0.95, P=0.89).

**** Insert Table 1 near here ****

Classification scores manifested 96% (intensity) and 54% (frequency) of concordances between the videotape protocol and the wearable system. Wilcoxon tests evidenced significant differences between F_{NEW} and F_{ST} (P<0.001). The maximum difference between the indexes was always of one point of the scoring scale.

F_{NEW} and F_{ST} appeared linearly distributed, with a $R^2=0.91$. All the trials were included in the Confidence Interval (CI) identified in the Bland-Altman plot, with a range of 0.11 tics/s [CI=(-0.02; 0.09)].

DISCUSSION

The aim of the current study was to develop a wearable actigraph and a tic-detecting algorithm for the automatic quantification of motor alterations caused by the Tourette Syndrome. A preliminary analysis about the potentialities and applicability in a clinical environment was carried out: the system performances were compared with the standard video-based analysis.

The sensitivity, specificity and accuracy of the tic-detection process were around 80%. This result is appreciable because it means that the algorithm was robust to differentiate tics from normal movements. It provided a good compromise between false positives and false negatives, despite a few limiting factors. (i) To our knowledge, this study represents the first proposal for an automatic and quantitative assessment of motor tics. Hence, due to the lack of standards, we designed the steps and set the parameters of the algorithm to optimize all the performance indicators (ACC, SPE and SEN) over the whole population. (ii) The system had not specifically devised to detect and evaluate the facial tics, which are common manifestations in TS. At this initial stage we preferred to focus onto the gross motor manifestations of TS, leaving to future developments the identification of a wider spectrum of TS-related events, including the less easily detectable. (iii) The wearable device consisted of a single sensor which was fixed on the trunk of the patient, independently from the possible different locations and intensities of tic bursts. The idea behind this was to test an unobtrusive, simple device that could work despite the inter- and intra-individual variability in tic expression and in different motor conditions (standing and walking). Tic acceleration patterns measured at the trunk are an indirect measure of the actual ones occurring in different body areas. Their characteristics (intensity and frequency content) are subject- and symptoms-dependent. Nevertheless, the system was able to detect pathological events far from the device location, provided that the phenomenon was intense enough.

The system accuracy, specificity and sensitivity may be improved by developing more effective detection algorithms. However, the definition of minimal requirements for detecting performances is needed. The limit of acceptance may be set only after a thorough debate among experts in the field [21], on the base of experience and of clinical goals/risks, like it has been done for other well established clinical tests (e.g. insomnia and sleep apnoea). In addition, spread sensors positioning and/or individual calibration procedures may enhance the efficacy of the detection system. Applying additional sensors may help in identifying peripheral low-intensity events, but may threaten the ecology and unobtrusiveness of the test. Patient-specific pre-calibration may solve the issue of variability in tic expression within and between individuals. This could allow the tic recognition algorithm to work in an optimal condition for each TS subject, with a proper intensity scale modulation and a specific frequency filtering.

In the classification stage, tic-intensity scores consistently matched the ones provided through standard clinical methods. Frequency indexes, F_{NEW} and F_{ST}, often manifested a one-point difference on the scoring scale, but the two measures had a significant correlation, when expressed in tics/s. This may relate to the definition of the 0-4 scoring intervals of F_{NEW}, which referred to the gold-standard ones. However, the sensitivity of the two measuring systems is very likely to be different, because our method is based on automatic recognition while the traditional one relies on human abilities. Hence, a point for future improvements may be the adjustments of F_{NEW} scoring ranges, which may be settled by exchanging data and experiences from different labs/research groups on larger sample sizes.

The least square regression analysis and the Bland-Altman plot evidenced that the new method had a good accuracy in terms of tic-frequency estimation. Though, as mentioned above, the automatic detection procedure appeared to have a slight bias in estimating the number of tics occurred. The

confidence interval identified by the B-A analysis proved that the difference between F_{NEW} and F_{ST} may be wide enough to produce a scoring error of one point of the frequency scale.

In this study we focused only on tic-intensity and tic-frequency domain. However, the collected acceleration signal processing are likely to retain additional information, that may be realistically used to assess other tic features: anatomical distribution of tic events, that may be evaluated by analysing acceleration patterns distinctly along the three directional sensing axis and/or by adding/distributing further sensors (e.g. accelerometers, gyroscopes) on the body; complexity and muscular involvement, that may be correlated to duration and shape of accelerometric waveforms produced by tic events. These features may be evaluated in future, to provide a more thorough and accurate characterization of motor disorders associated to TS.

In conclusion, a novel wearable measurement system was proposed for the automatic recognition and classification of motor tic features due to the Tourette Syndrome. The proposed paradigm has a high potentiality, using such device as a simple, portable and cost-effective support system. In addition, new indexes were defined to quantify tic frequency and intensity according to the medical scales which are currently in use and represent the standard for clinical evaluations. The system may be further improved in terms of: accuracy, extending its capabilities also to facial and vocal tics; focus on the individual; use during other types of motor activities. Though, the preliminary results are promising, and the method may emerge as a conventional tool for both clinical investigations and home-environment/long-term monitoring of TS motor disorders. To reach this goal and to create the bases for the definition of new standards, efforts should be spent to increase the trial database and to stimulate the collaboration and the exchange of experiences between different research groups.

ACKNOWLEDGEMENTS

AUTHORS' ROLES

1. Research project: A. Conception, B. Organization, C. Execution;

2. Statistical Analysis: A. Design, B. Execution, C. Review and Critique;

3. Manuscript: A. Writing of the first draft, B. Review and Critique;

Bernabei: 1, 2, 3.

Preatoni: 1A, 1B, 2A, 2C, 3.

Mendez: 2A, 2C, 3B.

Piccini: 1A, 1B, 2C, 3B.

Porta: 1A, 1C, 3B.

Andreoni: 1A, 1B, 2C, 3B.

Full Financial Disclosures of all Authors for the Past Year

Michel Bernabei

Stock Ownership in medically-related fields	None	Intellectual Property Rights	None
Consultancies	None	Expert Testimony	None
Advisory Boards	None	Employment	None
Partnerships	None	Contracts	None
Honoraria	None	Royalties	None
Grants	None	Other	None

Ezio Preatoni

Stock Ownership in medically-related fields	None	Intellectual Property Rights	None
Consultancies	None	Expert Testimony	None
Advisory Boards	None	Employment	Research Fellow at Politecnico di Milano
Partnerships	None	Contracts	None
Honoraria	None	Royalties	None
Grants	None	Other	None

Martin Mendez

Stock Ownership in medically-related fields	None	Intellectual Property Rights	None
Consultancies	None	Expert Testimony	None
Advisory Boards	None	Employment	Research Fellow at Politecnico di Milano
Partnerships	None	Contracts	None
Honoraria	None	Royalties	None
Grants	None	Other	None

Luca Piccini

Stock Ownership in medically-related fields	None	Intellectual Property Rights	None
Consultancies	None	Expert Testimony	None
Advisory Boards	None	Employment	Research Fellow at Politecnico di Milano
Partnerships	None	Contracts	None
Honoraria	None	Royalties	None
Grants	None	Other	None

Mauro Porta

Stock Ownership in medically-related fields	None	Intellectual Property Rights	None
Consultancies	None	Expert Testimony	None
Advisory Boards	None	Employment	None

Partnerships	None	Contracts	None
Honoraria	None	Royalties	None
Grants	None	Other	None

Giuseppe Andreoni

Stock Ownership in medically-related fields	None	Intellectual Property Rights	None
Consultancies	None	Expert Testimony	None
Advisory Boards	None	Employment	Assistant Professor at Politecnico di Milano
Partnerships	None	Contracts	None
Honoraria	None	Royalties	None
Grants	None	Other	None

REFERENCES

1. Tourette Syndrome Classification Study Group. Definitions and classification of tic disorders. Arch Neurol 1993;50:1013–1016.

2. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. Washington, DC: American Psychiatric Association; 1994.

3. Shaprio AK, Shapiro ES, Young JG, Feinberg TE. Measurement in tic disorders. In: Shapiro AK, Shapiro ES, Young JG, Feinberg TE, editors. Gilles de la Tourette syndrome. 2nd ed. New York: Raven Press; 1988. p 452–480.

4. Goetz CG, Tanner CM, Wilson RS, Carroll S, Como PG, Shannon KM. Clonidine and Gilles de la Tourette's Syndrome: double-blind study using objective rating methods. Ann Neurol 1987;21:307– 310.

5. Goetz CG, Leurgans S, Chmura TA, Home alone: methods to maximize tic expression for objective videotape assessments in Gilles de la Tourette syndrome. Mov Disord 2001;16(4):693–697.

6. Goetz CG, Pappert EJ, Louis ED, Raman R, Leurgans S, Advantages of a modified scoring method for the rush video-based tic rating scale. Mov Disord 1999;14(3):502–506.

7. Leckman JF, Riddle MA, Hardin MT, et al. The Yale Global Tic Severity Scale: initial testing of a clinician-rated scale of tic severity. J Am Acad Child Adolesc Psychiatry 1989;28:566–573.

8. Frontera W. The importance of technology in rehabilitation. IEEE Eng Med Biol Mag 2003; 22(3):25.

9. Bonato P, Mork PJ, Sherrill DM, Westgaard RH. Data Mining of Motor Patterns Recorded with Wearable Technology. IEEE Eng Med Biol Mag 2003;22(3):110-119.

10. Aminian K, Robert P, Buchser EE, Hayoz D, Depairon M, Rutschmann B, Physical activity monitoring based on accelerometry: validation and comparison with video observation. Med Biol Eng Comput 1999;37:304-308.

11. Aminian K and Najafi B. Capturing human motion using body-fixed sensors: outdoor measurement and clinical applications. Computer Animation and Virtual Worlds 2004;15:79–94.

12. Patel S, Hughes R, Huggins N, Standaert D, Growdon J, Dy J, Bonato P. Using Wearable Sensors to Predict the Severity of Symptoms and Motor Complications in Late Stage Parkinson's Disease. Conf Proc IEEE Eng Med Biol Soc. 2008;2008:3686-9.

13. Moore ST, MacDougall HG, Gracies JM, Cohen HS, Ondo WG. Long-term monitoring of gait in Parkinson's disease. Gait & Posture 2007;26:200–207.

14. Bonato P. Advances in wearable technology and applications in physical medicine and rehabilitation. J Neuroeng Rehab 2005;2:2.

15. Stein J. Wearable sensor technology for functional assessment after stroke. IEEE Eng Med Biol Mag 2003;22(3):26-27.

16. Mathie M, Coster A, Lovell N, and Celler B. Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement. Physiol Meas 2004;25:R1–R20.

17. Goetz CG, Tanner CM, Wilson RS, Shannon KM. A rating scale for Gilles de la Tourette's syndrome: description, reliability, and validity data. Neurology 1987;37:1542–1544.

18. Goetz CG, Stebbins GT, Thelen JA. Talipexole and adult Gilles de la Tourette's syndrome: double-blind, placebo-controlled clinical trial. Mov Disord 1994; 9:315–317.

19. Jin F, Sattar F, Goh DY. An acoustical respiratory phase segmentation algorithm using genetic approach. Med Biol Eng Comput 2009;47(9):941-53.

20. Bland JM, Altman DJ, Comparing methods of measurement: why plotting difference against standard method is misleading. The Lancet 1995; 346:1085-1087.

21. Westgard JO, Hunt MR. Points of care in using statistics in method-comparison studies. Clin Chem ;44(11):2240-2242.
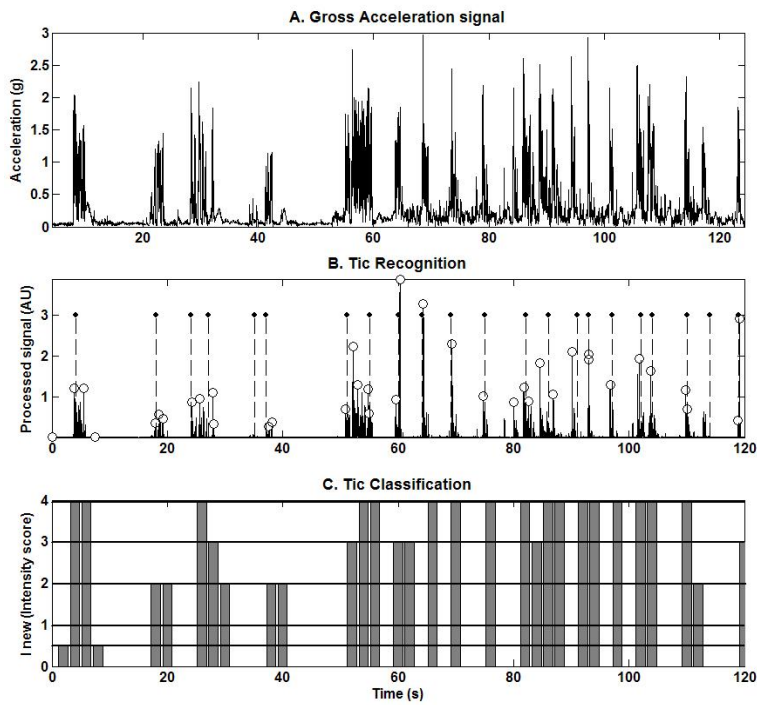
FIGURES AND TABLES

Figure 1. Example of the data processing workflow concerning a single clinical trial on a TS subject. A) Acceleration signal acquired through the wearable device, the subject was in standing position for the first 60 s and was walking in the last 60 s. B) The processed signal after NEO filtering and thresholding: (o) indicates the tic events identified by the algorithm; (•) represents tics actually occurred and identified by the physician. C) Classification of the intensity of each recognized event in a 0-4 rating score.

Table 1. Sensitivity (SEN), specificity (SPE) and accuracy (ACC) percentage performances (mean and standard deviation) of the automatic tic-recognition algorithm. SEN, SPE and ACC were averaged over all testing trials and reported for both frontal (FRONT) and dorsal (BACK) device positioning. M-W column reports P-values of the Mann-Witney tests between FRONT and BACK conditions.

		FRONT	**BACK**	**M-W**
SEN	$SEN = \dfrac{TP}{TP + FN}$	80.8 (8.5)	78.9 (9.2)	0.77
SPE	$SPE = \dfrac{TN}{TN + FP}$	75.8 (17.3)	78.3 (12.8)	0.95
ACC	$ACC = \dfrac{TP + TN}{TP + FN + TN + FP}$	80.5 (12.2)	80.4 (7.7)	0.89

TP= true positives; TN= true negatives; FP= false positives; FN= false negatives.

SUPPLEMENTARY MATERIAL

The main aim of the signal processing adopted in this paper was to automatically recognize the onset of tic events in the background of accelerations produced by normal movements. Digitized 3-axial acceleration signals were the inputs of the algorithm, whose output consisted of an array of instants of tic-occurrence. These time-events may be easily represented on the time-scale and compared with the tic set manually identified by the physician through videotape inspection. This comparison allowed to assess the algorithm performances in terms of sensitivity and specificity of tic recognition.

Tic Recognition

Recognition of tics from video recordings and from acceleration signals were completely independent and were eventually compared only after they had performed their identification of pathological events. This means that the experienced physician who carried out the patient's assessment according to Goetz's methodology [1] was completely blind for acceleration data. Similarly, the algorithm for the detection of pathological spikes was completely automatic and did not rely on visual observation of recordings.

The algorithm consisted of 3 subsequent steps.

i) Non-linear median filtering (NMF). NMF was applied to remove the signal "baseline", which may be related to posture, non-pathological slow movements and breathing. The use of NMF implies the proper selection of Ω (order of the median filter) which may affect the sensitivity of tic recognition.

ii) NEO (Non-linear Energy Operator) filtering [2]. It represents a frequency-independent energy estimator that has been already used to detect noisy spikes in biological signal processing. NEO involves a thresholding procedure, whose aim is to distinguish between spikes due to tics and spikes due to fast non-pathological movements. NEO's thresholding involves the choice of

another critic parameter, *m*, that influences the degree at which accelerations were assigned to possible pathological events

iii) Adaptive thesholding. We assumed that the spectral distribution of the acceleration signal tends to spread over a larger range of frequencies when the subjects turns from standing to walking (Figure 1). The variance of the spectrum may thus be seen as a movement dependent factor. This quantity was used to tune the adaptive thresholding involved in the NEO filter and to provide a self-calibration of the algorithm that relies on the amount of motor activity.
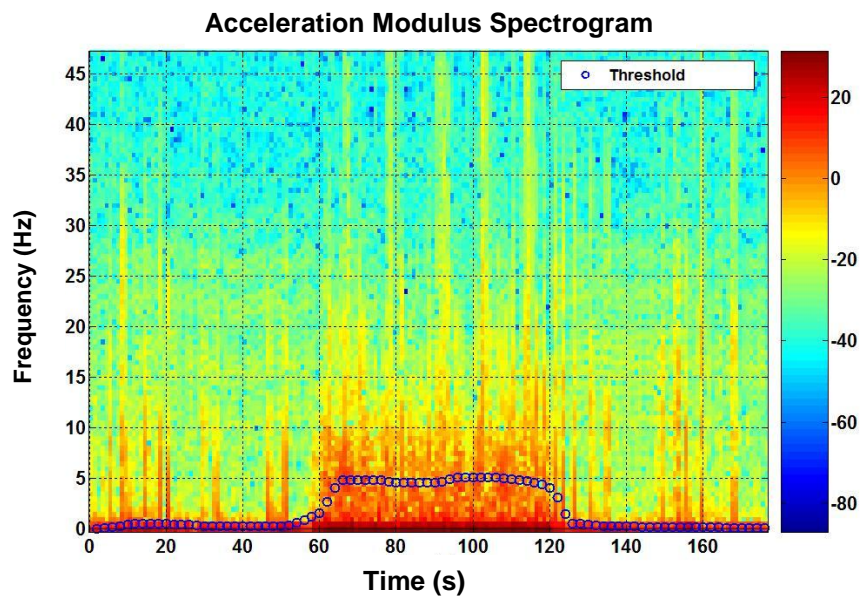

Figure 1. Example of spectrogram from analyzed data. Frequency spectral density of the acceleration signal over 512-samples windows is reported through a colorimetric scale. 0-60 and 120-180 time intervals refer to a standing session while during the 60-120 period the subject was walking. The spectrum variance over time (blue dotted line) may distinguish between the motor tasks.

Therefore we continuously determined the threshold (*T*) over 2-seconds windows. *T* depended on: the NEO algorithm, and the variance of the spectrum of acceleration signal over time, in a *"the higher the variance, the higher the limit"* non-linear fashion. Equations [1] and [2] report a more detailed description of the formula.

$$\textbf{[1]:} \quad T = \frac{C}{N} \times \overset{N}{\underset{n=1}{\text{å}}} (NEO(a(n))) \Theta (f(n))$$

- *C* is a scaling factor depending on the spectral variance;

- *N* is the number of samples; *NEO(a(n))* is the NEO filtering of acceleration;

- *Θ* denotes convolution and *f(n)* is a 6-point Bartlett window.

$$\textbf{[2]:} \quad C = m \cdot \text{var}_{spectrum}$$

- $var_{spectrum}$ is the variance of the distribution of the spectrum,

- *m* is a coefficient to be determined. Particularly, *m* is one of the two parameters on which the sensitivity/specificity/accuracy analysis was based.

Hence, every acceleration peak identified and isolated through the NEO algorithm and the 2-sec windowing that goes over the time-dependent value of the threshold *T* is chosen as a probable tic event and consequently classified through the intensity scale (fig. 2):

I_{NEW} (g=9.81 m/s):

5. barely perceptible: $0 \leq I_{NEW} < 0.5$;

6. visible: $0.5 \leq I_{NEW} < 1$;

7. some problem: $1 \leq I_{NEW} < 1.5$;

8. impaired function: $1.5 \leq I_{NEW} < 2$;

9. no function: $I_{NEW} \geq 2$.

The algorithm provides both a local information about the intensity of each recognized tic and an overall severity classification based on intensity/frequency of tics. The I_{NEW}/F_{NEW} indexes were calculated through a mean over local intensity values and through a (number of tics)/(time period) ratio respectively.
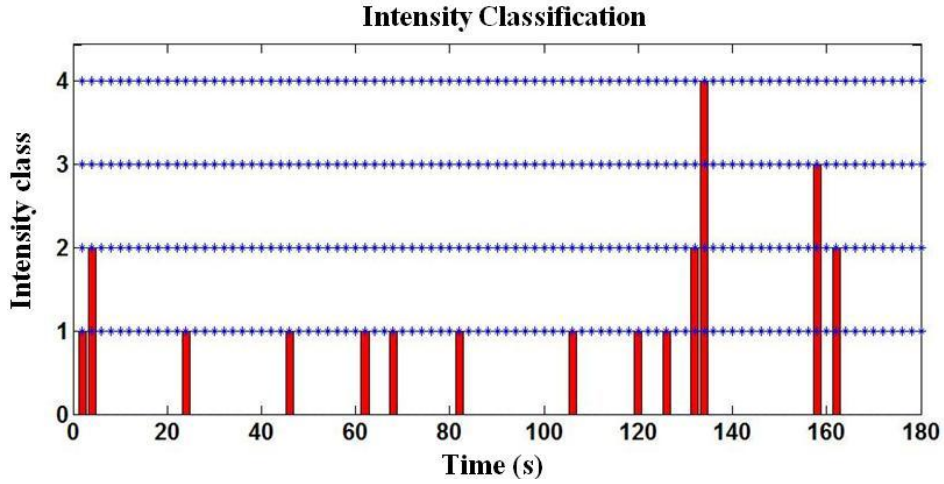
Figure 2. Example of the recognition/classification process. In a 180-seconds period the algorithm found 15 tic events and classified almost all of them in the first class (visible), with a maximum in the fourth class (no function).

Performance Evaluation

The identification of tics does not rely only on the determination of the threshold (hence on *m*), but also on the order of the median filter that was used to subtract the baseline from the initial acceleration waveform (hence also on parameter *Ω*).

We performed a parameter-dependent evaluation of the algorithm based on the correspondence of actual tics (recognized through the videotape) and the estimated ones. Therefore we set up our sensitivity/specificity/accuracy analysis both on *m* and on *Ω*, following these steps:

1) we estimated specificity and sensitivity as functions of the two parameters ($SPE(m,Ω)$, $SEN(m,Ω)$), for each subject (Figure 3);

2) we averaged results from single subjects over the whole population;

3) we plotted a graph, where the average dependence of *SEN* and *SPE* upon *m* and *Ω* is represented by 3D surfaces (Figure 4). This representation may be reconducted to the information provided by the ROC curve in a mono-dimensional domain (i.e. only 1 parameter considered, Figure 5). Therefore we used it to determine the values of *m* and *Ω*

that represent the "best compromise" in terms of detection performances, and that were used for the comparison of the automatic algorithm with the gold-standard.
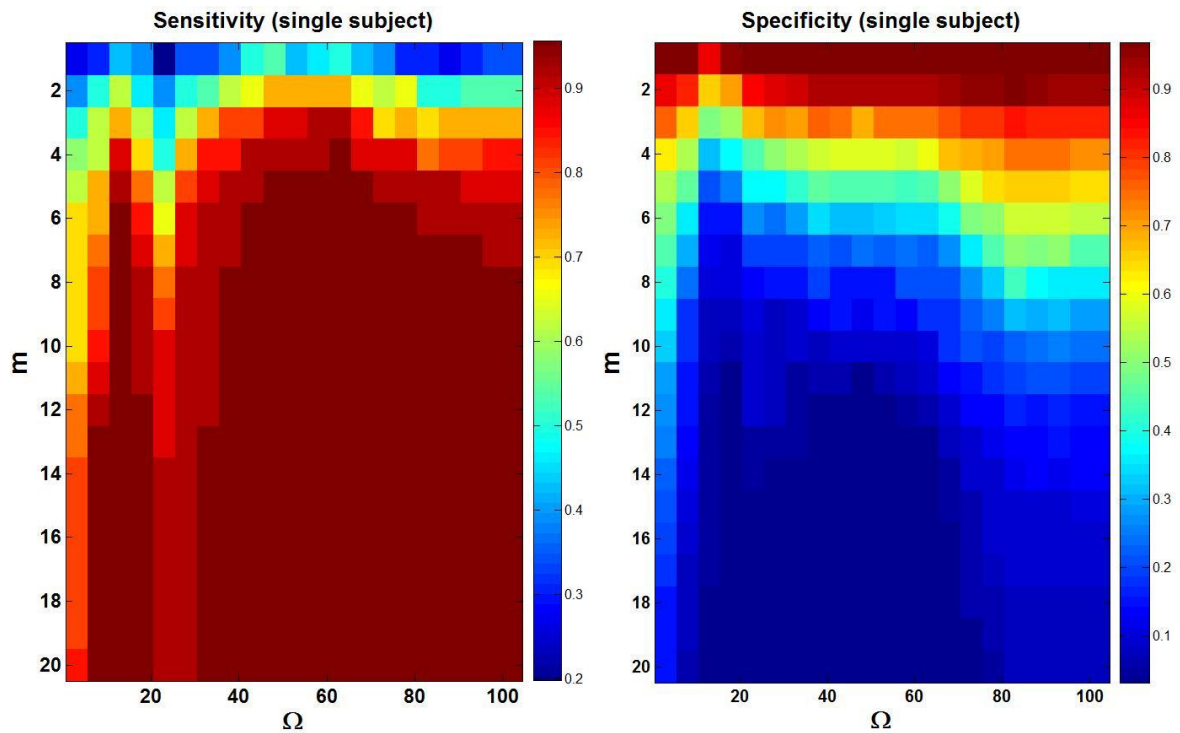

Figure 3. Performance indexes (i.e. sensitivity/specificity percentages) in dependence parameters *m* and *Ω* are represented through colorimetric scales. The more the color is near dark red the more the algorithm was able to recognize actual tics.
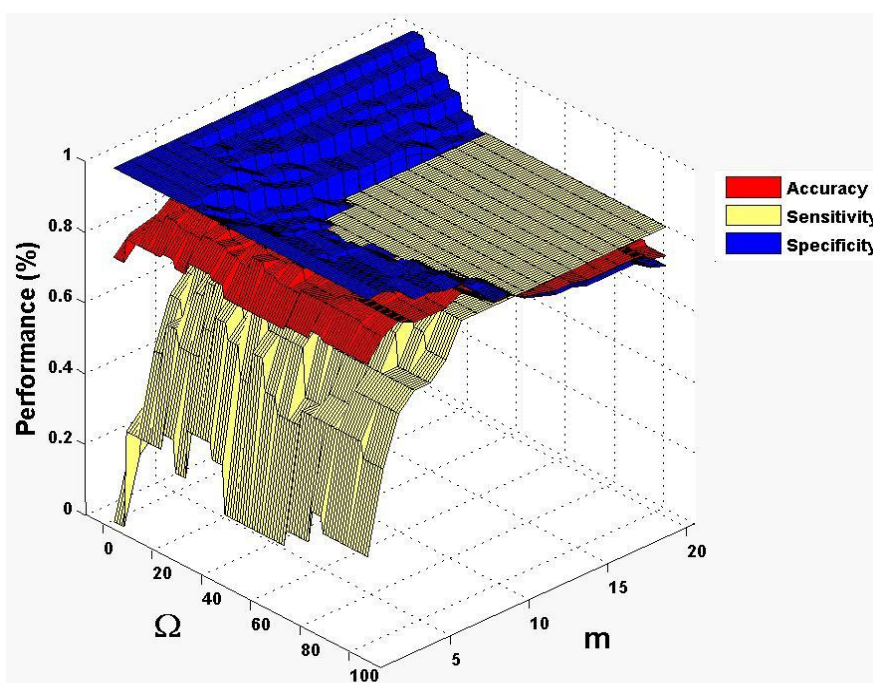
Figure 4. Sensitivity, specificity and accuracy surfaces in the 3D space. The performance planes intersect at a certain value of *m* (8-10) that may ensures a good compromise between sensitivity/specificity.

It may be observed that parameter *Ω* has, on average, less influence on performance than parameter *m*. The choice of *m* may thus represent an important factor of improvement in detection potentialities, in particular referring to the individual. Tuning *m* individually through a specifically designed calibration phase at the beginning of the acquisition may be a way to enhance the automatic tic-recognition performances.
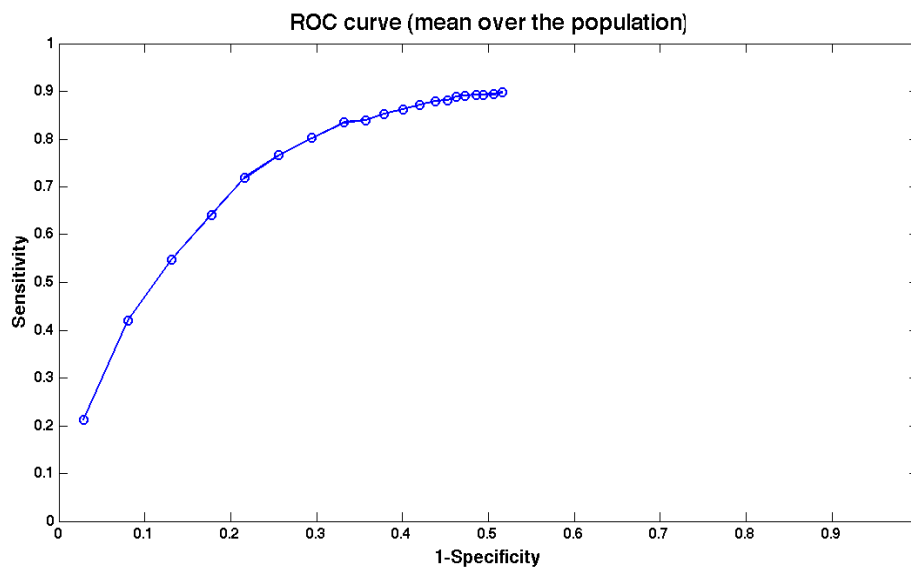

Figure 5. ROC curve averaged over the whole population, relative to the variation of parameter *m*, when *Ω* is made constant. It may be seen as a one-dimensional approximation of a multidimensional information obtained averaging the 3D surface along its less variable direction (i.e. the *Ω* axis).

References

22. Goetz CG, Pappert EJ, Louis ED, Raman R, Leurgans S, Advantages of a modified scoring method for the rush video-based tic rating scale. Mov Disord 1999;14(3):502–506.

23. Jin F, Sattar F, Goh DY. An acoustical respiratory phase segmentation algorithm using genetic approach. Med Biol Eng Comput 2009;47(9):941-53.