Clinical assessment of freezing of gait in Parkinson's disease from computer-generated animation

Tiffany R. Morris, Catherine Cho, Valentina Dilda, James M. Shine, Sharon L. Naismith, Simon J.G. Lewis, Steven T. Moore

ABSTRACT

The current 'gold standard' for clinical evaluation of freezing of gait (FOG) in Parkinson's disease (PD) is determination of the number of FOG episodes from video by independent raters. We have previously described a robust technique for objective FOG assessment from lower-limb acceleration. However, there is no existing method for validation of autonomous FOG measures in the absence of video documentation. In this study we compared the results of clinical evaluation of FOG from computer-generated animations (derived from body-mounted inertial sensors) during a timed up and go test with the 'gold standard' of clinical video assessment, utilizing a cohort of 10 experienced raters from four PD centers. Agreement between the 10 clinical observers for scoring of FOG from computer animations was more robust for the relative duration of freeze events (percent time frozen; intraclass correlation coefficient of 0.65) than number of FOG episodes, and was comparable with clinical evaluation of the patient from video (intraclass correlation coefficient 0.73). This result suggests that percent time frozen should be considered (along with number of FOG events) to better convey FOG severity. The ability of clinical observers to quantify FOG from computer-generated animation derived from lower-limb motion data provides a potential approach to validation of accelerometry-based FOG identification outside of the clinic.

1. Introduction

Freezing of gait (FOG), a paroxysmal locomotor block, is a debilitating symptom of Parkinson's disease (PD) associated with an increased risk of falls and early nursing home placement [1]. Clinical management of FOG is limited in large part by the difficult nature of assessing its severity, particularly in a community setting. Current approaches utilize subjective reports from patients or caregivers (either as a simple rating scale or FOG questionnaire), but our recent study of a cohort of PD patients with self-reported FOG symptoms demonstrated that scores from subjective FOG questionnaires did not correlate with actual freeze severity when walking [2]. The de facto 'gold standard' of FOG assessment is clinical evaluation (either directly or from video) to determine the number of FOG episodes [5,6], but this subjective approach also has limitations, as evidenced by a moderate inter-rater agreement (intraclass correlation coefficient of 0.6) in our recent study of 10 experienced movement disorder specialists assessing FOG from video [3]. Clinical management of FOG, as well as the evaluation of new targeted interventions, would benefit from the development of objective, standardized FOG measures capable of monitoring this debilitating symptom in a community setting.

We have described an objective technique for robust identification of FOG events from lower limb accelerometry, which has achieved strong agreement with clinical assessment in the laboratory [3,4]. Our ultimate goal is to extend accelerometry-based FOG assessment to community monitoring. However, a critical barrier to validation of such an approach is the current inability to verify FOG in the absence of clinical observation. Although accelerometry has proven viable in a controlled clinical setting [3,4], validation of autonomous FOG monitoring outside of the clinic requires an intermediate step: a means to allow clinical assessment in the absence of video documentation. To this end we developed a technique for visualizing objective lower body motion data of ambulating patients. The aim of this study was to assess the utility of clinical FOG assessment from these 3D computer animations by comparison with the current 'gold standard' of assessment from video of the actual patient. Our hypothesis was

* Corresponding author at: Mount Sinai School of Medicine, Department of Neurology, Box 1052, 1 Gustave Levy Place, New York, NY 10029, USA.
Tel.: +1 212 241 1943; fax: +1 212 831 1610.
E-mail address: steven.moore@mssm.edu (S.T. Moore).

0966-6325/$ – see front matter © 2013 Published by Elsevier B.V.

that computer animation of lower-limb motion generated from
body-mounted inertial sensors, in particular the high-frequency
(3–8 Hz) ‘trembling’ associated with FOG [3,4], would provide
sufficient information to enable accurate clinical evaluation of
freezing. Such a finding would provide a much needed means of
validating objective FOG monitoring outside of the clinic, by
facilitating a comparison of clinical FOG evaluation from com-
puter-generated animations with accelerometry-based FOG identifi-
cation [3,4], both derived from ambulatory motion sensors.

2. Methods

2.1. Patients

Ten PD patients (6 male and 4 female) with a clinical history of FOG were
recruited from the Parkinson’s Disease Research Clinic at the Brain and Mind
Research Institute, University of Sydney. Patients were assessed in the practically
defined ‘off’ state following overnight withdrawal of dopaminergic therapy. Two
patients were being treated with subthalamic nucleus deep brain stimulation (DBS),
which was turned off for 60 min prior to testing. Patient characteristics have been
described previously [3], and are summarized here: mean age 67.7 [SD 6.6],
disease duration 11.2 years [SD 6.7], Hoehn and Yahr stage 2.5 [SD 0.3], UPDRS –
Section III [7] 38.4 [SD 10.7]. None of the patients described any increase in
freezing behavior following the administration of their usual dopaminergic therapy.

2.2. Protocol

Subjects were instrumented with seven inertial measurement units (IMUs –
Xsens MTx, Enschede, Netherlands) secured to the back (approximately L2),
the lateral aspect of each thigh and shank, and the superior aspect of each mid-foot,
with elastized straps. The IMUs were small (38 mm × 53 mm × 21 mm; 30 g)
and did not interfere with natural movement. As previously described, patients
performed timed up-and-go (TUG) tasks to provoke FOG on a standardized 5-m
course, while being recorded on a digital video camera from a consistent
target point for clinical analysis [2,3]. Fourteen TUG tests were performed
with a total duration of 9 min 37 s (four of the ten subjects who performed the
TUG task quickly, or who exhibited minimal FOG, were asked to perform a
second trial).

2.3. Animations

During testing each IMU acquired triaxial linear acceleration, angular velocity
and readings of the Earth’s magnetic field; transmitted wirelessly to a computer at
a sample rate of 50 Hz. Synchronization of video and accelerometer recordings
was performed prior to data collection by alignment of the video camera and data-
acquisition computer clocks. Data was processed post hoc to determine the
orientation of each sensor array in space using a commercial sensor fusion
algorithm (MT Manager, Xsens, Enschede, Netherlands). IMUs do not provide
absolute position data therefore, to generate an animation of patient gait during
the TUG task we created a virtual humanoid lower body with two robotic chains
(right and left leg) based on the ‘Nancy’ model [8] using the LabVIEW Robotics
Module (National Instruments, Austin, TX) (Fig. 1). Each IMU generated a matrix
representing the orientation of the sensor (and associated limb segment) in 3D
space, which was referenced to a neutral (upright) position determined from a
pre-TUG calibration (320 ms during quiet stance):

\[ M_i = M_{\text{f}} \times N_i^{-1} \]

where \( N_i \) is the orientation matrix representing the neutral position for sensor \( i \), \( m_i \)
is the original (uncalibrated) orientation matrix representing the current
orientation of the IMU in space, and \( M_i \) represents the orientation of this body
segment with respect to the neutral position \( N_i \). After aligning all sensor orientations
with respect to neutral, joint angles (\( f \)) for the left and right leg were calculated [9]
in a downward cascade from the orientation of the distal sensor with
respect to the proximal sensor for the hip (thigh relative to back), knee (shank
relative to thigh) and ankle (shank relative to foot) (Fig. 2).

\[ J_{\text{hip}} = M_{\text{f}} \times M_{\text{shank}}^{-1} \]
\[ J_{\text{knee}} = M_{\text{shank}} \times M_{\text{ankle}}^{-1} \]
\[ J_{\text{ankle}} = M_{\text{f}} \times M_{\text{ankle}}^{-1} \]

The yaw orientation of the back sensor determined heading (the direction in which
the patient’s avatar was facing; Fig. 1). The resultant joint angles, limited to the
range of the biomechanically possible (hip 121° flexion to −20° extension; knee
−143° flexion to 5° extension; ankle 13° flexion to −56° extension [10]), were used
to drive the robotic chains of the humanoid model, and an animation was created
from each image at a rate of 50 frames per second (Fig. 1 and supplementary video).

2.4. Clinical assessment of FOG

Clinicians experienced in evaluation of freezing (n = 10) were recruited from four
Parkinson’s disease centers (Mount Sinai, the Bronx VA, and Beth Israel in New
York City; the University of Sydney) to independently assess the videos and animations
for FOG (the results for rater agreement from video have been previously
presented [3]). Each rater reviewed a total of 32 videos, which comprised 14 unique videos
of ambulating patients and 14 corresponding animations, and 4 repeated video/animation
pairs to assess intra-rater (within each observer) reliability. The 32 video/animations
were presented in a random order (with no two observers having the same sequence) such that clinicians were blinded as to which video a
particular animation was associated with. Raters were not informed that some
patients had more than one video/animation trial or that four of the video/animation
pairs were presented twice (all observers viewed the same four repeat trials).
The avatars were identical for each animated trial; therefore observers had
no visual cue to patient identification apart from lower-body motion. The
frequency and relative duration of freezing episodes for each video and animation
were quantified with a FOG tagging program [2,3]. Each rater used their best clinical
judgment to identify FOG episodes, tagging the onset of a freeze by pressing the ‘T’
key and holding the key for the duration of each event. The program saved the
clinical ratings as a binary signal with a baseline of zero (no freeze) and a value of 1
indicating a freeze event, from which the number of FOG events and percent time
frozen (the cumulative duration of all FOG episodes divided by the total duration
of the walking task) were calculated [2,3].

Fig. 1. A sequence of video frames from a patient performing the TUG task and the corresponding computer-generated animation (see also supplementary video).

2.5. Statistical analysis

Video and animation from the 14 unique TUG trials from the 10 patients were used to assess inter-rater (across the 10 observers) reliability; the four repeat videos were used to assess intra-rater reliability. The reliability of the quantification of the number of FOG episodes and percent time frozen from video and animation was calculated between the 10 clinical observers, within each observer for the four repeat assessments, and within each observer for video versus animation, using the intraclass correlation coefficient (ICC(1,1)) [11]. In addition, the mean and 95% confidence interval (CI) of the number of FOG events and percent time frozen from the 10 raters was calculated for each of the 14 animated TUG trials (Fig. 2).

3. Results

The reliability of clinical ratings from video has been described previously [3]; here we compare the results of clinical assessment from 3D animations generated by movement data with the video results. Quantification of FOG from animation was in general agreement with video assessment; mean number of FOG events reported in a single trial (Fig. 3A) varied from 0.9 (CI 0.9) to 4.3 (CI 6.1) [0 (CI 0) to 6 (CI 3.0) from video]; percent time frozen ranged from 4.0 (CI 5.1) to 86.2 (CI 6.2) [0 (CI 0) to 79.8 (CI 6.0) from video (Fig. 3B)]. Inter-rater (between the 10 observers) reliability from animation was weak for number of FOG events (ICC = 0.35), but considerably higher for percent time frozen (ICC = 0.65) (video results were 0.63 and 0.73, respectively [3]). Similarly, agreement between the video and animation ratings for each observer were much stronger for percent time frozen (mean ICC 0.67 [CI 0.07]) than number of FOG events (mean ICC 0.27 [CI 0.20]). A paired t-test confirmed (p < 0.0001) that observers scored more FOG episodes (both number and relative duration) from animation relative to video (Fig. 3).

Intra-rater (within observer) reliability for scoring of the four repeat animations was strong (mean ICC 0.71 [CI 0.19]) for percent time frozen, but moderate for number of FOG events (mean ICC 0.47 [CI 0.20]), comparable to results from video assessment (mean ICC 0.71 and 0.44, respectively [3]).

A repeated measures ANOVA was conducted to confirm the validity of our ICC results. Validity is deemed to be suspect if the intra-subject variability is not significant; between subjects ANOVA for both percent time frozen and number of freezes were highly significant (p < 0.001).

Comparison of Clinical Measures of FOG from Animation and Video

Fig. 3. (A) Clinicians scoring for number of FOG events (mean and CI) from animation (solid trace) and video (dashed trace) for the 14 trials. (B) Clinicians scoring for percent time frozen (mean and CI) from animation (solid trace) and video (dashed trace) for the 14 trials.

4. Discussion

Clearly video is a simpler method of clinical evaluation of FOG. The aim of this study was to validate a technique to allow clinical assessment of FOG when video or direct observation is not available, allowing objective measures of FOG to be compared to clinical observation during activities of daily living. The results demonstrate the viability of clinical assessment of FOG from computer-generated animations derived from lower-body acceleration data, and provide a potential approach to validation of accelerometry-based FOG identification [3,4] outside of the clinic [12,13].

With the exception of trial #4 (Fig. 3), raters tended to see one more FOG event in the animation than the corresponding video. This may have been due to motion artifacts in the computer-generated representation, or perhaps even due to animations exhibiting more nuance in high frequency leg tremor, allowing clinicians to discern more distinct FOG events. Either way, the fact that experienced observers exhibit a large inter and intra-rater variability in number of FOG [3] suggests little impact on overall assessment. Simply counting the number of discrete FOG episodes is unreliable [3]; calculation of the relative duration of freezing (percent time frozen) minimizes differences in inter-rater preference for registering multiple or concatenated FOG episodes, and is a much more robust metric of FOG severity [3]. In this study agreement between the 10 clinical raters was higher for percent time frozen than number of FOG, consistent with our previous study of clinical evaluation from video [3]. Similarly, within-rater reliability was much stronger for percent time frozen from animation (and video [3]). These results were likely due to a disparity in information content between the two metrics. With the exception of trial #4, there was little variation in scoring for number of FOG episodes across raters for both animation and video assessment, with a positive result of typically 1 or 2 FOG events per TUG task. In contrast, percent time frozen exhibited a considerable range (4–86% from animation, 0–80% from video); simply counting the number of FOG events did not convey the severity of FOG. For example, the animations of TUG trials #8 and #13 were scored similarly in terms of number of FOG events (n = 2), whereas percent time frozen in trial #13 (68%) was more than double that of trial #8 (33%). The current ‘gold standard’ focuses on determining the number of FOG events [5,6]; the results of this and previous studies [2,3] suggest that the relative duration of freeze events (percent time frozen) should also be considered to better represent the extent of FOG impairment.

Video, in addition to being simpler to implement, conveys significantly more information than lower-limb animation, from upper body motion to the facial expressions of both patient and attending clinician. Disadvantages include the purely subjective nature of assessment, the necessary presence of a video operator or clinician, and potential privacy issues (particularly if filmed outside of the clinic). Drawbacks of our computer-generated animation approach are complexity and the potential for introduction of artifactual motion; these limitations will likely diminish as motion capture techniques inevitably improve over time. A compelling advantage of animation is the linkage between conclusions based on clinical observation and physical movement data. Computer-generated animations provide not only a 3D pictorial representation of patient movement through space, but also the underlying kinematics, allowing direct comparison of objective and subjective measures of FOG. This is exemplified by the strong correlation between accelerometer measures of high-frequency (3–8 Hz) leg movement and clinical identification of FOG episodes [3,4]. Taken together, our results suggest that this ‘trembling’ of the lower limbs, as well as step length and cadence, are robust clinical features in freezing of gait.

Acknowledgement

The authors gratefully acknowledge the patients and clinicians who participated in this study.

Conflict of interest statement

The authors declare they have no conflict of interest.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.gaitpost.2012.12.011.

References