KEYSTROKE ANALYTICS FOR NON-INVASIVE DIAGNOSIS OF NEURODEGENERATIVE DISEASE

Andrew D. Ellington
Tim Riedel
Dan Winkler
Emily Knight

2015
UT CID Report #1513

This UT CID research was supported in part by the following organizations:

identity.utexas.edu
KEYSTROKE ANALYTICS FOR NON-INVASIVE DIAGNOSIS OF NEURODEGENERATIVE DISEASE

Executive Summary

We sought to use the temporal dynamics of keyboarding during natural computer typing as an indicator of identity and health status. We first developed novel keystroke logging software in Python. We then analyzed the hold times (time between keydown and keyup for each key) and flight times (time between subsequent key presses) for two healthy individuals and found that 1) hold times and flight times differ significantly (p < 0.001) between these individuals, and 2) hold times for an individual are consistent across different times of day and different days. We then acquired typing data from patients with Parkinson’s disease (PD) (n=16) and from elderly controls (n=15) who typed for 15 minutes each, copying the same passage of text. We extracted several features based on the temporal dynamics of each subject’s typing. Two features in particular - a measure of consistency of hold time probabilities as a function of experiment time, and the degree of asymmetry in hold times between hands - reasonably separated the patients from the controls (classification accuracy: 65%). For patients, we also looked for correlations between each feature and the clinical score (UPDRS III) indicating the degree of severity of motor symptoms. None of the features we tested correlated with PD motor symptom severity. This study represents a proof-of-principle that keyboarding has promise for both identity and diagnostic purposes. In future work we hope to modify our keylogging software to be web-based and to crowd-source data collection in order to acquire a much larger data set.

Introduction

Identity can mean a variety of things, but should include the notion of how disease plays into identity. Our health and healthcare are definitive parts of who we are, and increasingly this is true online. We proposed to develop a method whereby the onset of neurodegenerative diseases can be detected virtually, via the interaction of the user with the most common ‘transducer’ of information, the keyboard (or touch screen). By better understanding the role of health and disease in this process, we could also pave the way to a more personal understanding of how individuals interact with their devices, a key goal for the Center for Identity.

Earlier detection of neurodegenerative diseases is of paramount importance because the best treatments for these diseases only slow the progression; no treatments can reverse the damage wrought to the brain. We propose to use the relatively well-known technique of logging and analysis of keyboard strokes to provide insights into the health status of individual patients with Parkinson’s disease. Keystroke logging would in essence become an early warning monitor that can allow users to seek care at a point where intervention is most effective (and cost-effective).
The analyses we proposed differs from standard keyboard logging in that they will focus on both idiosyncrasies of keyboarding and context-dependent changes in keyboarding.

We seek to improve upon existing diagnostic techniques and therapeutic monitoring for Parkinson's disease (PD), which would result in new technology that improves service quality and outcomes for patients. Specifically, we seek to:

1. Develop software that logs sequence and timing of keyboard strokes,
2. Curate a database of normal keyboarding behavior by healthy individuals to serve as a baseline for subtle shifts that may signify onset of disease, and
3. Develop a statistical machine-learning software pipeline that accurately classifies PD from healthy controls.

**Background and Hypotheses**

It is estimated that 1 million Americans have Parkinson's disease. The >65 year-old segment of the population is amongst the fastest growing, including in terms of computer use. More importantly, as newer generations that are already computer-savvy age, they will encounter the same issues but be far more willing to incorporate onboard diagnostics into their daily lives.

PD is a progressive, degenerative disorder of the central nervous system that manifests with a variety of motor symptoms such as tremor, rigidity, slowness or freezing of movements, and gait problems, and therefore has a tremendously deleterious impact on quality of life. There is recent evidence that some treatments may help slow further progression of the disease. Thus, timely diagnosis is critical for ensuring patients have the best possible therapeutic outcome. We propose to develop algorithms for early detection of PD by mining keyboarding data.

The motor symptoms of PD have several attributes that are likely to be extractable from keyboarding data. First, the disease affects the left and right hemispheres of the brain separately and therefore exhibits asymmetric symptom patterns between the left and right hands. Second, irregularities in the timing of repetitive, alternating two-finger key strikes correlate with severity of PD and therefore show promise as a diagnostic tool (Tavares et al., 2005Printy et al., 2014). We propose to capitalize upon a richer data source from keyboarding with all 10 fingers for a more detailed assessment of PD. Third, tremor, freezing of movement, and rigidity manifest in different patterns during repetitive movements and should affect typing in ways that differ from healthy subjects. Finally, since keyboarding is not a purely motor task but requires cognitive function, the types of errors produced by PD patients are expected to differ from those of healthy controls given the mild cognitive deficits associated with PD.

A limited literature also exists on the general use of keystroke analysis to infer information about a user's cognitive state. While keystroke analysis has been used for user authentication and computer security for decades (for review see Gunetti et al., 2005), most methods involve building a profile of a user's typical typing pattern based on a short, fixed set of keystrokes, with the sole purpose of
differentiating that pattern from the patterns of unauthorized users. Changes in the typing pattern, either from changes in emotional state, arousal, or physical health status have been considered undesirable noise that makes recognizing a user of a computer system more difficult. In contrast, Giancardo and colleagues capitalized on this type of variability to show that typing patterns, specifically the duration and consistency of the time that keys are held down, may be used to infer if individual users are drowsy or alert (Giancardo et al, 2015). Epp and colleagues (2011) also demonstrated that particular aspects of users’ typing patterns may be used to build models that detect users’ emotional states. Since the symptoms of PD also involve cognitive and psychomotor changes, we hypothesize that keystroke dynamics may be used to detect these changes during normal computer use with consumer hardware. This would allow the creation of a PD diagnostic tool that is minimally intrusive, inexpensive and automatic.

Before exploring whether keyboarding could be used diagnostically, we first sought to confirm that keyboarding can be used for identity purposes.

Hypothesis
1a): Keyboarding can be used to distinguish between individuals.
1b.) Keyboarding patterns that differentiate between individuals should persist through different times of day.

Hypothesis 2: Patterns of flight times between pairs of keys should differ between patients and age-matched controls. Specifically:
2a.) Asymmetry in hold times between left and right hands should be greater for patients with Parkinson’s disease than for controls.
2b.) Hold times should show greater variance (lower self-similarity) or PD than for controls during a 15 minute typing task, since PD patients experience progressive bradykinesia.

Methods

*Human Subjects*

For the “identity” portion of the project (Hypothesis 1), two U.T. undergraduates from the Freshman Research Initiative explored natural keystroke logging. They logged keystrokes on their personal laptop computers (both running Microsoft Windows 8) while carrying out their typical daily typing. The students logged their typing for 4 days before analyzing the logs.

Patients with Parkinson’s Disease were approached for consent to participate in the study at an outpatient neurology clinic in Austin, TX. Healthy control participants were recruited from among family members accompanying patients at the neurology clinic and from senior living communities in the Austin area. All participants gave informed consent according to guidelines of the UT Austin Institutional Review Board. The age of the groups were 69.1 +/- 10.9 for PD patients (mean, std, n =
16, 7 women) and 70.6 +/- 9.9 (n = 15, 12 women), for controls. Additional details about participant demographics and disease profile are outlined in Tables 1 and 2.

Patients were first administered the Unified Parkinson's Disease Rating Scale (Fahn et al 1987, Goetz et al 2008) by their neurologist. As part of the research study, subjects were then administered the Montreal Cognitive Assessment (MOCA), a screening tool for mild cognitive impairment and dementia, as well as the Beck Depression Inventory (BDI) by members of the research team. All participants self-reported a limited medical history including neurological and orthopedic history and current medications, as well as their typical computer use habits, typical typing frequency and style.

Table 1. Patient Demographics

<table>
<thead>
<tr>
<th>Subject ID</th>
<th>Sex / Age</th>
<th>MOCA</th>
<th>BDI</th>
<th>Touch-typer</th>
<th>UPDRS III</th>
<th>Disease Duration (Years)</th>
<th>Worse Side</th>
</tr>
</thead>
<tbody>
<tr>
<td>K001</td>
<td>F / 76.6</td>
<td>27</td>
<td>12</td>
<td>Y</td>
<td>29</td>
<td>5</td>
<td>Left</td>
</tr>
<tr>
<td>K003</td>
<td>M / 69.6</td>
<td>29</td>
<td>3</td>
<td>Y</td>
<td>13</td>
<td>6</td>
<td>Right</td>
</tr>
<tr>
<td>K009</td>
<td>F / 70.4</td>
<td>28</td>
<td>6</td>
<td>N</td>
<td>18</td>
<td>7</td>
<td>Left</td>
</tr>
<tr>
<td>K012</td>
<td>M / 68.5</td>
<td>26</td>
<td>11</td>
<td>N</td>
<td>11</td>
<td>4</td>
<td>Left</td>
</tr>
<tr>
<td>K013</td>
<td>M / 60.5</td>
<td>26</td>
<td>7</td>
<td>N</td>
<td>39</td>
<td>8</td>
<td>Left</td>
</tr>
<tr>
<td>K014</td>
<td>M / 72.3</td>
<td>23</td>
<td>10</td>
<td>N</td>
<td>31</td>
<td>2</td>
<td>Right</td>
</tr>
<tr>
<td>K015</td>
<td>M / 63.3</td>
<td>29</td>
<td>4</td>
<td>N</td>
<td>16</td>
<td>5</td>
<td>Right</td>
</tr>
<tr>
<td>K016</td>
<td>M / 78.9</td>
<td>26</td>
<td>1</td>
<td>N</td>
<td>33</td>
<td>3</td>
<td>Right</td>
</tr>
<tr>
<td>K017</td>
<td>F / 67.1</td>
<td>30</td>
<td>8</td>
<td>N</td>
<td>12</td>
<td>6</td>
<td>Left</td>
</tr>
<tr>
<td>K018</td>
<td>F / 84.6</td>
<td>25</td>
<td>22</td>
<td>N</td>
<td>38</td>
<td>5</td>
<td>Left</td>
</tr>
<tr>
<td>K020</td>
<td>M / 49.6</td>
<td>27</td>
<td>24</td>
<td>Y</td>
<td>32</td>
<td>5</td>
<td>Right</td>
</tr>
<tr>
<td>K023</td>
<td>F / 62.2</td>
<td>24</td>
<td>8</td>
<td>Y</td>
<td>24</td>
<td>9</td>
<td>Left</td>
</tr>
<tr>
<td>K029</td>
<td>F / 81.3</td>
<td>29</td>
<td>14</td>
<td>Y</td>
<td>18</td>
<td>3</td>
<td>Right</td>
</tr>
<tr>
<td>K030</td>
<td>M / 86.9</td>
<td>23</td>
<td>2</td>
<td>N</td>
<td>35</td>
<td>7</td>
<td>Right</td>
</tr>
<tr>
<td>K031</td>
<td>F / 51.2</td>
<td>30</td>
<td>1</td>
<td>Y</td>
<td>25</td>
<td>2</td>
<td>Right</td>
</tr>
<tr>
<td>Subject ID</td>
<td>Sex / Age</td>
<td>MOCA</td>
<td>BDI</td>
<td>Touch-typer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-----------</td>
<td>------</td>
<td>-----</td>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K002</td>
<td>F / 50.4</td>
<td>30</td>
<td>9</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K004</td>
<td>F / 59.3</td>
<td>28</td>
<td>2</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K019</td>
<td>F / 61.1</td>
<td>26</td>
<td>13</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K021</td>
<td>M / 70.3</td>
<td>-</td>
<td>-</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K022</td>
<td>F / 68.2</td>
<td>30</td>
<td>7</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K024</td>
<td>F / 83.4</td>
<td>25</td>
<td>2</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K025</td>
<td>F / 83.7</td>
<td>n/a</td>
<td>4</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K026</td>
<td>M / 81.2</td>
<td>29</td>
<td>2</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K028</td>
<td>M / 81.4</td>
<td>30</td>
<td>2</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K033</td>
<td>F / 69.1</td>
<td>26</td>
<td>2</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K034</td>
<td>M / 71.0</td>
<td>27</td>
<td>3</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K035</td>
<td>F / 63.0</td>
<td>29</td>
<td>0</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K036</td>
<td>F / 60.9</td>
<td>29</td>
<td>4</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K037</td>
<td>F / 78.7</td>
<td>21</td>
<td>10</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K038</td>
<td>F / 69.8</td>
<td>26</td>
<td>1</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K039</td>
<td>F / 77.5</td>
<td>27</td>
<td>16</td>
<td>Y</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Demographic Overview of Control Subjects

Although we tried our best to enroll controls who matched the patients, the group differed on several demographics including cognitive scores and depression scores. Montreal Cognitive Assessment (MOCA) scores were lower for patients than for controls, indicating greater impairment in the PD group. Beck Depression Inventory (BDI) scores were higher for patients than for controls, indicating greater depression in the PD group. These differences were to be expected, given that Parkinson’s disease has both cognitive and affective symptoms. It is important to note that any differences seen in keyboarding data between the groups might not be due to motor impairment, but to the variety of symptoms that affect people with Parkinson’s disease. MOCA
scores for patients were 26.8 +/- 2.3 (mean +/- std) and controls 27.4 +/- 2.5. BDI scores for patients were 8.8 +/- 6.8 (mean +/- std) and for controls were: 5.1 +/- 4.8. Figure 1 illustrates the distributions of scores for these two metrics.

Figure 1. Statistics on demographics comparison of patients versus controls

**Keyboarding Task**

Participants were asked to complete a typing task on a 2011 model MacBook Air with an external wireless mouse. The experimental task involved 15 minutes of natural typing sample obtained by copying the content from the Wikipedia entry for salt into a Word document (https://en.wikipedia.org/wiki/Salt). Participants were instructed to copy as much of the text as possible in the allotted time, while typing naturally. This task was modeled from a similar task used by Giancardo and colleagues (2015) to detect psychomotor change, although the text was fixed rather than selected by the user, because fixed text typically increases classification accuracy (Gunetti et al., 2005; Epp et al., 2011).

**Data Processing**

Timestamped logs of each keystroke were obtained with custom Python software. An upper bound for the temporal variance of timestamps was established as being between 5ms and 10ms. This was
verified through operating system-generated keystrokes and using audio recordings of physical keystrokes. Custom data processing scripts were written in R, Python, and Matlab.

Several features were extracted from the processed data. For each logged key stroke, the hold time (time from when a key is pressed to when it is released) was calculated. Flight times (time from the release of a key to the press of the next key) were calculated for consecutive key presses. These features are depicted in Figure 2.

Figure 2. Schematic drawing for definitions of flight times and hold times, from Giancardo et al. (2015).
Figure 3. Raw key hold time data from example patient (top, subject K009) and control subject (bottom, subject K033).

Following the methods of Giancardo et al (2015), two features were extracted that summarize the predominant hold time and its robustness over time during the 15 minute typing task. In summary, the hold time data were binned into 2.5 second epochs, and histograms representing hold time probabilities were generated for each 2.5 epoch. The Key Hold Time Matrix, K, shows the probability density of hold times as a function of experiment time (Figs 4, 5). A parameter (Q = 10) set the minimum number of data points in a column of K necessary to include that column in further calculations; this effectively excluded points in time when the participant paused in typing, such as during the first 10 seconds of keystroke recording for subject K009 (Fig 3 top). Bin widths for the histograms (comprising the columns of matrix K, Figs 4,5) were 10 msec, and hold time values greater than 500 msec were relocated to the final bin. Kp is the Key Hold Time Evolution Peak, calculated as the peak hold hold from each column of K, and Ks is the Key Hold Time Evolution Self Similarity, calculated as the sum of the self-similarity matrix generated by each column of matrix K. We normalized Ks by dividing by the total number of time bins in the experiment so that total amount of typing time would not influence the calculation of self-similarity (Figs 4, 5).
Figure 4a. Key Hold Time Evolution Matrix, K, from example patient K009.

Columns in this matrix represent probability distributions (histograms) of hold times. Note the gap in the first 10 seconds of data collection. Statistics from this period and similar gaps are not included in subsequent analyses. Values closer to red indicate the peak hold time for a particular column representing a time epoch.
Figure 4b. Key Hold Time Evolution Matrix, K, from example control subject who does touch-type (K033).

In contrast to his previous subject, this subject has shorter hold times on average, and with less variance. This results in a lower Kp value (see Figure 10, subject K033).
Figure 5a. Key Hold Time Self-Similarity Matrix, $S$, from example patient who does not touch type (subject K009).

The self-similarity matrix, $S$, is generated from the Euclidean distance between all pairs of columns from matrix K. Values of 0 represent perfect self-similarity (hence, the blue values along the diagonal). Feature $K_s$ is generated from the sum of this matrix. The greater the amount of red in this matrix, the higher $K_s$ value, representing less self consistency of hold times over time, and the greater the amount of blue, the lower the $K_s$ value, indicating greater self consistency of hold times over time.

Example subject K009 (Fig 5a) has a higher $K_s$ feature than example subject K033 (Fig 5b).
Figure 5b. Key Hold Time Self-Similarity Matrix from example control subject who does touch-type (K033).

In a separate analysis, hold times were divided into two groups according to the side of the keyboard (and right vs. left hand, in the case of touch-typers) that the keys belonged to. After removing outliers, histograms were generated for the right and left hand keys separately. Kolmogorov-Smirnov distance between distributions of hold times between right and left hand was calculated for each subject, capturing a summary of the subject’s “Hand Asymmetry Index.”

Finally, flight times were calculated between all possible key pairs, and the top 23 pairs were selected by a simple thresholding of the sum of all subjects’ data. Median and standard deviation of flight times for each of these key pairs was calculated for each subject, yielding two 1 x 23 vectors for each subject.

Classifier Model

We tested several classifier models on the ability to correctly classify patients and controls on the basis of the two most promising features. Using Weka 3, open-source Java-based machine learning
software, we tried a logistic classifier, a random forest classifier, and a naive Bayes classifier with 10-fold cross-validation.

### Results

**Identity Keystroke Analysis**

We used R statistical software to compare hold times of any key pressed. Figure 6 shows box plots of hold times of Subject 1 in red and Subject 2 in blue. Morning (M), afternoon (A), and evening (N) logging sessions are listed left to right. These results suggest that hold times can be used to distinguish between individuals, and that hold times do not vary significantly as a function of the time of day.

Figure 6. Comparison of keystroke hold times for two healthy individuals. Morning (M), afternoon (A), and evening (N) hold times are also compared.
Figure 7. Comparison of Key Flight times and Hold times for two healthy volunteers.

Figure 7 indicates that both flight times and pause times show different median values when Subject 1 and 2 are typing naturally. These differences were confirmed with t-tests (p-values < 0.0001). The high significance of these simple analytics not only supports the distinguishing power
of natural logging on identity, but suggests that even with the variation we propose from health effects, it should still be possible to distinguish one typist from another.

Figure 8. Both subjects were logged for four days and hold times are compared for morning (red), afternoon (green), and evening (blue) typing sessions.

The characteristic hold times remained highly stable over 4 days of logging for Subject 1. Subject 2 showed more time-of-day variation, particularly in the morning (Fig 8, red), but at no time overlapped with Subject 1.

The frequency of backspace keystrokes for each subject was calculated and shows potential to be an identifying characteristic (Figure 9).
Figure 9. Backspace keystrokes frequency was normalized by total keystrokes over logging periods for each subject. Subject 1 uses the backspace key almost 2.5 times more frequently when typing naturally than Subject 2.

Feature Analyses by Group (Patient vs. Control)

To explore which features had the potential to cluster elderly typists according to group (PD vs. control), we generated scatter plots of individual features, color coded according to group membership.

Figure 10 depicts Hold Time Evolution Matrix features Kp and normalized Ks (Giancardo et al., 205). Blue values are PD patients and red values are controls.
The combination of features Kp and Ks norm do not appear to cluster according to subject group.

We next computed “hand asymmetry” to see whether this feature would separate patients from controls. We hypothesized that patients would have greater hand asymmetry than controls due to the Parkinson’s disease and the fact that it affects the two sides of the body on different time scales. We first computed hold times for the keys associated with the right and left side of the keyboard separately. We then plotted these distributions of hold times.

Figure 11 shows left and right hand hold time distributions from an example PD patient, K003, who is a touch-typist with significant hand asymmetry.
Figure 11. Distributions of hold times by hand for patient K003. The distribution of hold times for left hand keys is on the top, and the distribution of hold times for right hand keys is on the bottom.

Finally, we computed the Kolmogorov-Smirnov distance between the distributions for each subject, as an indicator of Hand Asymmetry.

Figure 12 is an examination of whether the degree of hand asymmetry separates according to subject groups. Blue data points are PD patients and red points are controls. Since Figure 10 suggests that Ks_norm may have some power to discriminate between subject groups, we plot both Ks_norm and Hand Asymmetry for each subject to see whether the subjects cluster according to group.
Figure 12 shows that there may be some separability of groups on the basis of both Hand Asymmetry and normalized Key Hold Time Self Similarity. As was predicted by our hypothesis, patients tend to have greater hand asymmetry than controls do; this is because Parkinson’s disease affects the different sides of the body at different rates. Since a self-similarity value of 0 indicates perfect self-similarity, or consistency of hold times as a function of experiment time, controls tend to have better hold time robustness. This is also consistent with the prediction that PD patients would experience progressive bradykinesia and therefore have less consistency in hold times as a function of time.

We then examined flight times between common key pairs (digraphs) as a potential feature space.

Figure 13 is a color plot of mode flight times for the most common key pairs. Color values represent milliseconds. Since a subsequent key press may be released before the previous key press is released, flight times can sometimes take on negative values. Each column represents a subject, and each row represents a digraph (pair of keys). Patient data is in the top subplot, and control data are on the bottom.
Figure 13. Peak flight times for most commonly occurring digraphs, by subject. The columns in the top row represent patients, and the columns in the bottom row represent controls.

By eye at least there is no clear pattern that emerges that distinguishes between patients and controls. However, it appears that non touch-typers have greater variance in digraph flight times than touch typers. To investigate this, we plotted flight times for the top 23 digraphs for the four different subject groups.
Figure 14. Flight times for top digraphs for patients (top row) and controls (bottom) separated into touch-typers (left) and non-touch-typers (right).

Figure 14 demonstrates that the variances of flight times is a better indicator of whether a subject is a touch-typist or not, rather than an indicator of patient versus control.

Classification Results

Although the data set was smaller than one would like for machine learning, we tested several models on the ability to correctly classify the 16 patients and 15 controls on the basis of the two most promising features: Ks norm, and hand asymmetry. We modeled a logistic classifier, a random forest classifier, and a naive Bayes classifier, each tested with 10-fold cross-validation. The naive Bayes was the best performing model, with 65% of the data correctly classified and 35% incorrectly classified. A summary of results are listed in Table 2.

<table>
<thead>
<tr>
<th>Model</th>
<th>Correctly Classified Instances</th>
<th>Incorrectly Classified Instances</th>
<th>True Positive Rate</th>
<th>False Positive Rate</th>
<th>Area under ROC Curve</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic</td>
<td>61.3%</td>
<td>38.7%</td>
<td>0.613</td>
<td>0.392</td>
<td>0.654</td>
</tr>
<tr>
<td>Random Forest</td>
<td>54.8%</td>
<td>45.2%</td>
<td>0.548</td>
<td>0.457</td>
<td>0.617</td>
</tr>
</tbody>
</table>
Correlation Analyses

For patients only, we then plotted each feature as a function of UPDRS III score which is an indicator of severity of motor symptoms.

Neither the Key Hold Time Evolution Peak feature (Kp), nor the normalized Key Hold Time Evolution Self Similarity feature (Ks norm) were correlated with the severity of Parkinson’s disease motor scores (Figure 15).

Figure 15. Key Hold Time Evolution Peak (Kp) as a function of UPDRS III (left), and normalized Key Hold Time Evolution Self Similarity (Ks) as a function of UPDRS III (right).
We next plotted hand asymmetry as a function of UPDRS III to see whether this feature scaled as a function of severity of motor symptoms for the patients. Figure 16 indicates that there is no such relationship between hand asymmetry and disease severity.

Figure 16. Hand asymmetry (hold times) as a function of UPDRS III.

A summary of correlation analyses between extracted features and the UPDRS III are outlined in Table 3.

Table 3. List of features and r-values and p-values for correlation with UPDRS III for patients only

<table>
<thead>
<tr>
<th>Feature</th>
<th>Correlation with UPDRS III (r-value)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key Hold Time Evolution Peak (Kp)</td>
<td>-0.11</td>
<td>0.69</td>
</tr>
<tr>
<td>Normalized Key Hold Time Evolution Self-Similarity (Ks Norm)</td>
<td>0.37</td>
<td>0.15</td>
</tr>
<tr>
<td>Hand Asymmetry Index (K-S Distance between right-side and left-side keys)</td>
<td>0.08</td>
<td>0.77</td>
</tr>
</tbody>
</table>
Discussion

We first confirmed that keystroke features could be used to differentiate between individuals, and we found that this approach could work even for data acquired at different times of day.

We then sought to use keystroke analysis for health diagnostics. We collected samples from both Parkinson’s patients and from elderly healthy volunteers who typed from the same sample text for 15 minutes. Our general approach to analyzing these data was two-fold. First, we extracted features that we hoped would cluster according to group membership - patient versus control. The idea here was to identify a set of features that could predict whether an unknown subject falls into the PD group or the control group. Second, we looked for features that would scale as a function of disease severity. We plot extracted features as a function of the UPDRS III and perform linear correlation. The underlying idea for this approach is that a feature or combination of features that align to disease disease severity would show feature values that predict disease onset.

Unfortunately, the features explored during this study did not correlate with the UPDRS III, the neurologist rating that indicates the motor severity of a patient's symptoms. However, there is some suggestion that a combination of features, specifically the peak hold time (Kp) from the key hold time evolution matrix, and the hand asymmetry, together may be able to separate patients from controls.

Conclusions and Next Steps

The results of the analyses from this pilot study demonstrate first that temporal dynamics of keyboarding has the potential to help identify individuals. Additionally, we demonstrated a proof-of-principle that temporal dynamics of keyboarding has the potential to help diagnose neurodegenerative disorders, at least in the case of Parkinson’s disease. Additional data and analyses are clearly needed. From our small cohort, we found that the variance of hold times within subject groups was at least as great as the variance between groups. Parkinson’s disease is a highly heterogeneous disorder that affects patients very differently. The features examined herein did not scale as a function of disease severity. However, there were two features based on key hold times that showed promise for clustering the data into PD patient vs. healthy control groups based on a naive Bayes classifier. This model should be further tested with a significantly larger data set. We also feel that the ability to detect changes in typing on the basis of keyboarding could be greatly improved by acquiring data longitudinally from patients as the disease progresses. Having baseline data from individuals would potentially allow us to use these methods to detect subtle shifts in an individual's typing pattern, and would likely help increase the statistical power of these methods.
References


Poewe, W., Rascol, O. Sampaio, C., Stebbins, G., et al. The Unified Parkinson’s Disease Rating Scale (UPDRS) Status and Recommendations. Movement Disorders, 2003;18(7);783-750.


For more information on Center for Identity research, resources and information, visit identity.utexas.edu.