

# Collecting and Leveraging Identity Cues with Keystroke Analysis (CLICKA)

Oliver Buckley

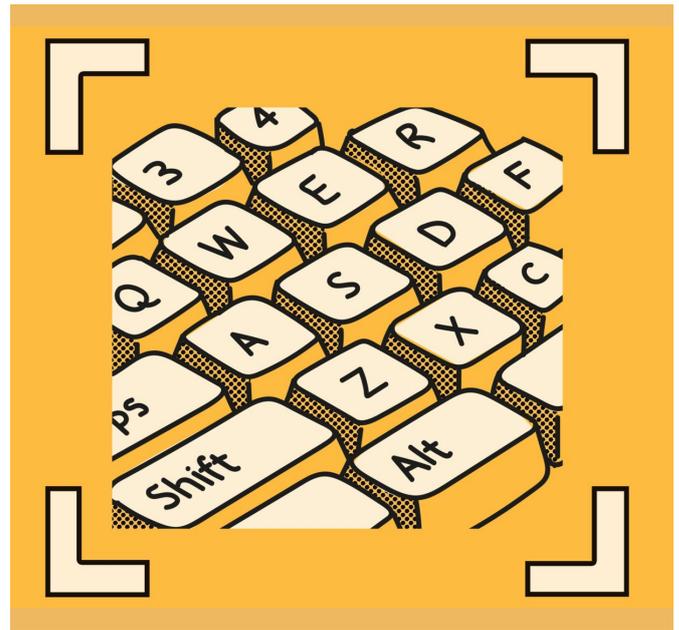
## EXECUTIVE SUMMARY

Keystroke dynamics is the analysis of how an individual uses a keyboard. These typing behaviours can be as uniquely identifiable as a person's handwriting or signature and this data can reveal identifying characteristics about an individual.

The analysis of the typing rhythm and cadence of a user can be used to identify an individual, while also providing information about the person sitting at the keyboard, which can include characteristics such as handedness, hand size, or typing style.

Whereas previous work largely focused on confirming the identity of an anonymous user, this work aimed to understand more about the individual using the device by determining the name and native language of an anonymous user, based on how they type.

The first experiment focused on determining the name of an anonymous user and collected typing samples from 84 users. The participants completed several typing exercises where the timing of each keystroke was recorded. The typing data were subdivided into substrings of two characters (bigrams), and the time between releasing the first key and pressing the second (the flight time) was calculated. The research hypothesis was that those bigrams with a greater familiarity with the user will have a discernibly higher ranking than those that are not as commonly used.



The research made use of machine learning classifiers to develop a model that is capable of a balanced accuracy prediction of approximately 70% of the bigrams in an anonymous user's name.

The second experiment aimed to predict the native language of an individual based on an analysis of their typing behaviours. The experiment collected data from 492 participants across five native languages (English, French, German, Spanish, and Italian) with around 100 people in each group. The participants were again required to complete typing exercises and this data was segmented into bigrams.

Again, machine learning classifiers were used to predict a user's native language. In the first instance, the research aimed to distinguish whether English was the user's native language (i.e. English versus French, Spanish,

## EXECUTIVE SUMMARY

### COLLECTING AND LEVERAGING IDENTITY CUES WITH KEYSTROKE ANALYSIS (CLICKA)

German, Italian) with a balanced accuracy of 71%, using the SVC classifier.

When predicting the native language of an individual, based on five languages, the approach achieved a balanced accuracy of 45%. While this offers significant room for improvement it does perform notably better than a random prediction.

The research established that users display repeatable and predictable typing behaviours based on familiar identity or linguistic data.

#### ABOUT THIS PROJECT

This Executive Summary was written by Oliver Buckley, School of Computing Sciences, University of East Anglia.

This summary details the key findings of work conducted by the CREST commissioned project *Collecting And Leveraging Identity Cues With Keystroke Analysis (CLICKA)*.

To read the Full Report and to view all the outputs from this project go to: [www.crestresearch.ac.uk/projects/clicka/](http://www.crestresearch.ac.uk/projects/clicka/)