Spoofing key-press latencies with a generative keystroke dynamics model

John V. Monaco Md Liakat Ali Charles C. Tappert

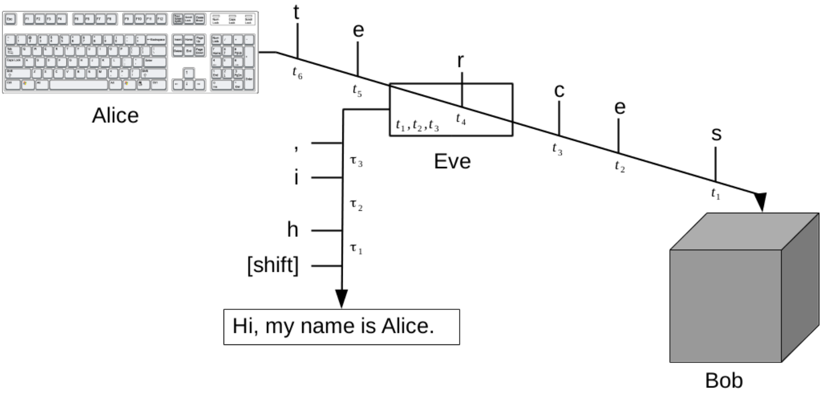
Pace University, NY

September 11, 2015

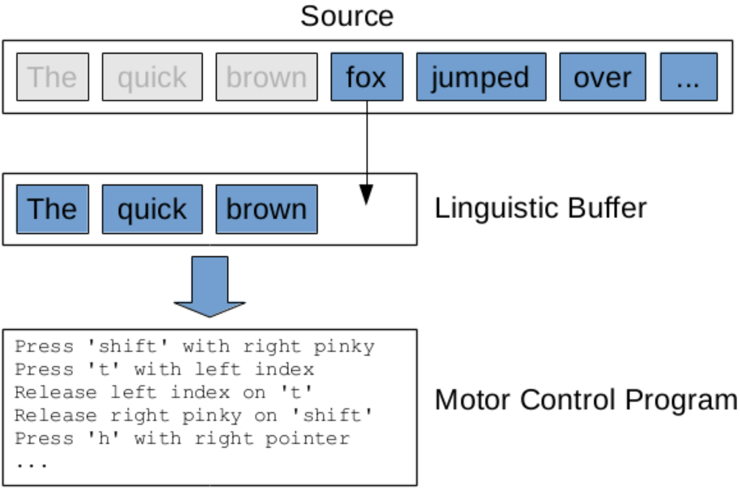
Outline

1. Introduction
2. Methodology
3. Experimental results
4. Conclusions

Scenario.


t
e
r
c
e
s
t_6
t_5
t_1, t_2, t_3
t_4
t_3
t_2
t_1
Alice
Eve
Bob
,
i
h
[shift]
τ_3
τ_2
τ_1
Hi, my name is Alice.

Typing behavior.



Source

The quick brown fox jumped over ...

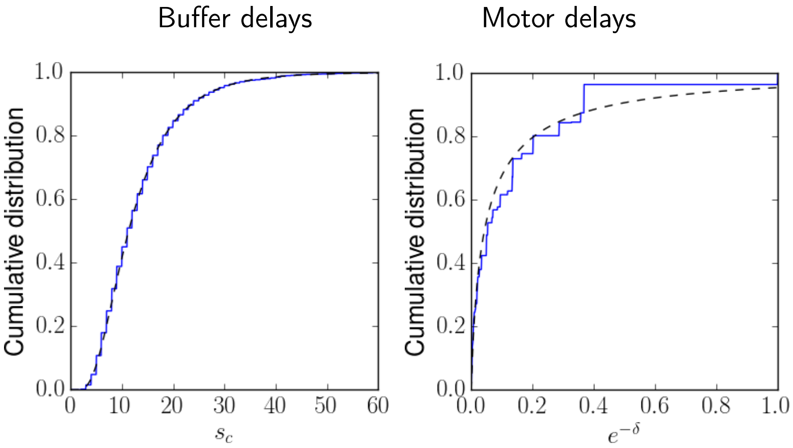
The quick brown

Linguistic Buffer

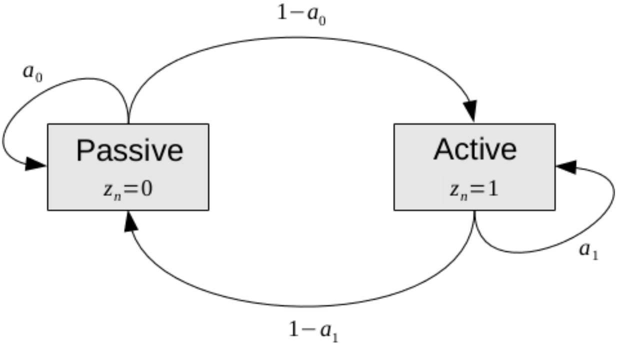
```
Press 'shift' with right pinky
Press 't' with left index
Release left index on 't'
Release right pinky on 'shift'
Press 'h' with right pointer
...
```

Motor Control Program

Predicted key-press latency distributions.



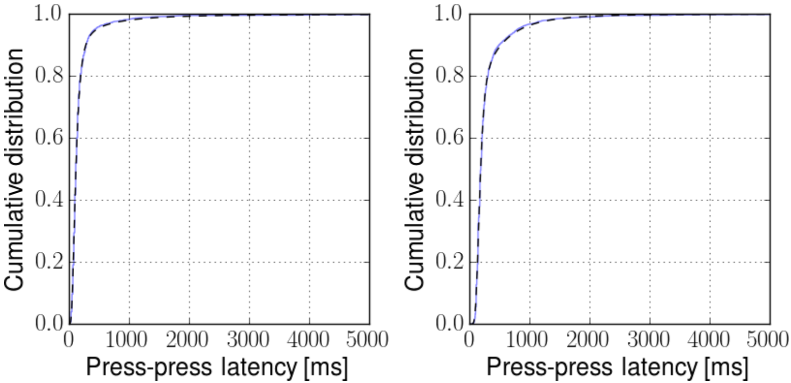
Two-state hidden Markov model.



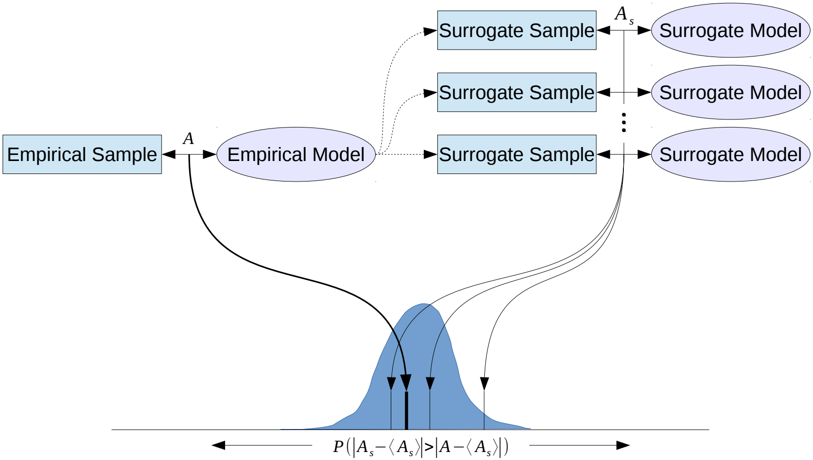
8 parameter model almost perfectly reproduces the empirical distribution of key-press latencies for every user

Empirical and model CDF.

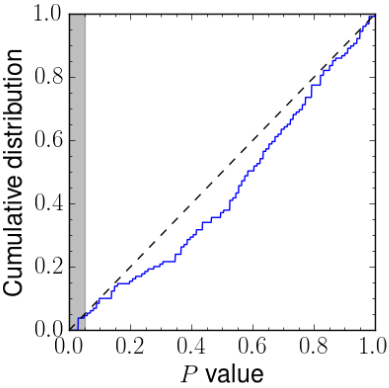
Empirical CDF (solid blue) and model CDF (dashed black) for 2 users



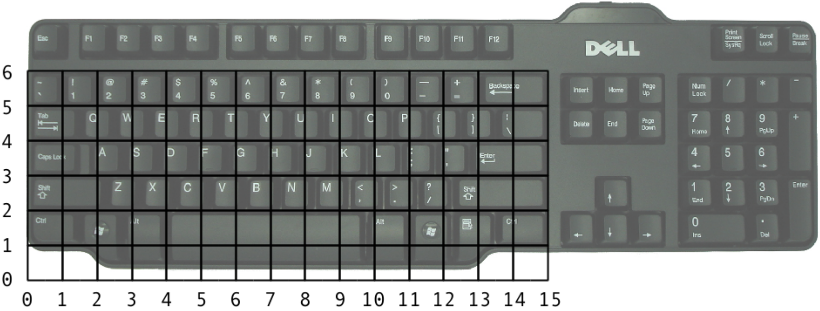
Goodness of fit test.



Goodness of fit test results.

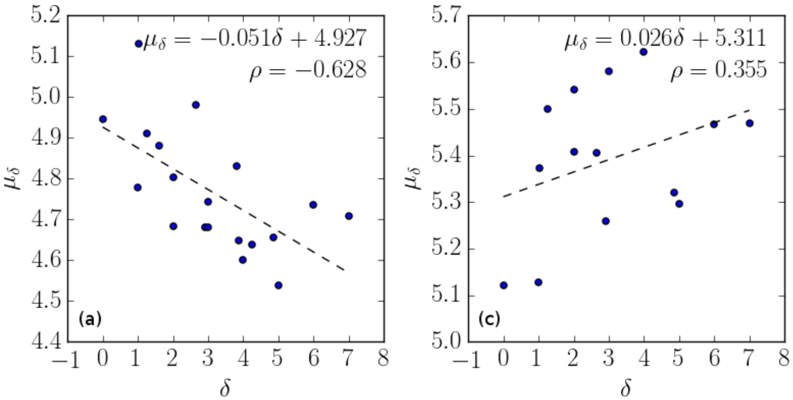


Keyboard coordinates.

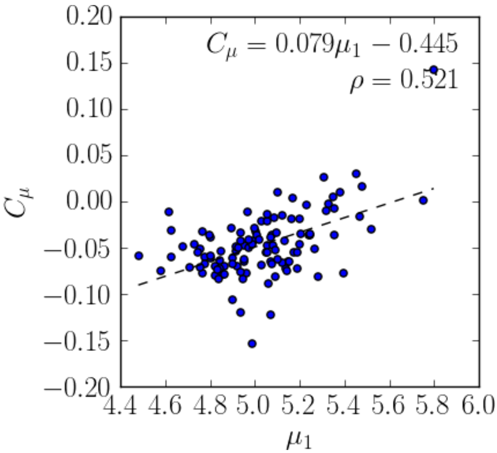


Scaling between latency and distance.

Log key-press latency vs. inter-key distance for fast and slow typists



Latency-distance slope vs. typing speed.



Spoofing procedure.

- Observe key-press latencies with missing key names
- Determine which latencies correspond to an active typing state using a 2-state HMM
- Use the latency inter-key distance scaling behavior to generate latencies for a predefined text

Recover the victim's typing behavior.

- Solve a system of equations to recover the expected key-press latencies for each unique inter-key distance in the predefined text

$$\mu_{\delta_i} - \mu_{\delta_j} = \frac{C_\mu}{\delta_i - \delta_j}$$

$$\sigma_{\delta_i} - \sigma_{\delta_j} = \frac{C_\sigma}{\delta_i - \delta_j}$$

$$\mu_s = \mu_1 = \sum w_\delta \mu_\delta$$

$$\sigma_s^2 = \sigma_1^2 = \sum w_\delta ((\mu_\delta - \mu_1)^2 + \sigma_\delta^2)$$

Empirical data.

- 129 users, 4 samples each
  - 751 ± 94 keystrokes per sample
- Key-press latency

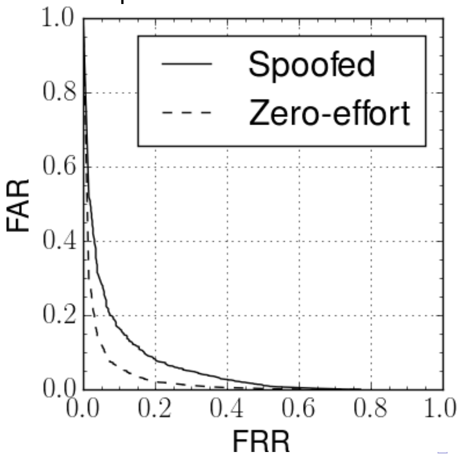
$$\tau_i = t_i - t_{i-1} \tag{1}$$

Experiment protocol.

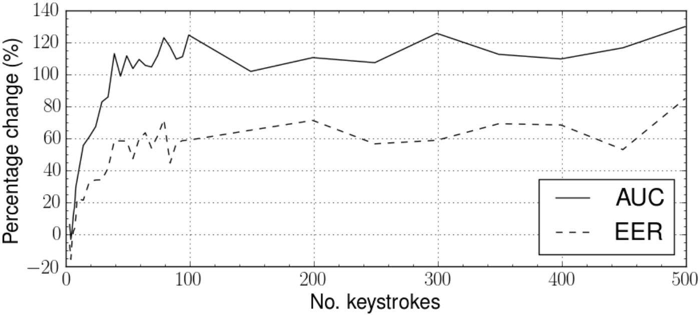
- Use the dichotomy classifier with key-press latency features
- Obtain zero-effort results in the usual way (authenticating every combination of users)
- Obtain spoofed results by observing the latencies with missing key names and generating a sample for the predefined text
- Stratified 4-fold cross validation

ROC curves for zero-effort and spoofed attacks.

Zero-effort: 7.5% EER
Spoofed: 12.9% EER



Relative increase in error over zero-effort.



Summary.

- With at least 50 observed keystrokes, the chance of success over a zero-effort attack doubles on average
- Worth exploring further?
  - Yes
- Next steps?
  - Model key-release times

Thank you.

Thank you