Relative Accuracy Measures for Stroke Gestures

Radu-Daniel Vatavu (University Stefan cel Mare of Suceava, vatavu@eed.usv.ro)

Lisa Anthony* (Computer & Information Science & Engineering (CISE), University of Florida, lanthony@cise.ufl.edu)

Jacob O. Wobbrock (Information School | DUB Group, University of Washington, wobbrock@uw.edu)



** work done while author was at UMBC Information Systems*

Universitatea Ştefan cel Mare Suceava

ICMI 2013

Relative Accuracy Measures for Stroke Gestures

Radu-Daniel Vatavu (University Stefan cel Mare of Suceava, vatavu@eed.usv.ro)

Lisa Anthony* (Computer & Information Science & Engineering (CISE), University of Florida, lanthony@cise.ufl.edu)

Jacob O. Wobbrock (Information School | DUB Group, University of Washington, wobbrock@uw.edu)



Authors of the ICMI 2012 Best Paper Award winning paper "Gestures as Point Clouds: A $P Recognizer for User Interface Prototypes"

** work done while author was at UMBC Information Systems*

Universitatea Ştefan cel Mare Suceava

ICMI 2013

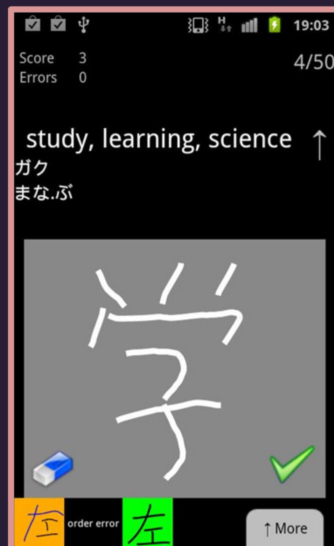
Outline

1. Motivation
2. The Relative Accuracy Measures
3. Results
4. Conclusion

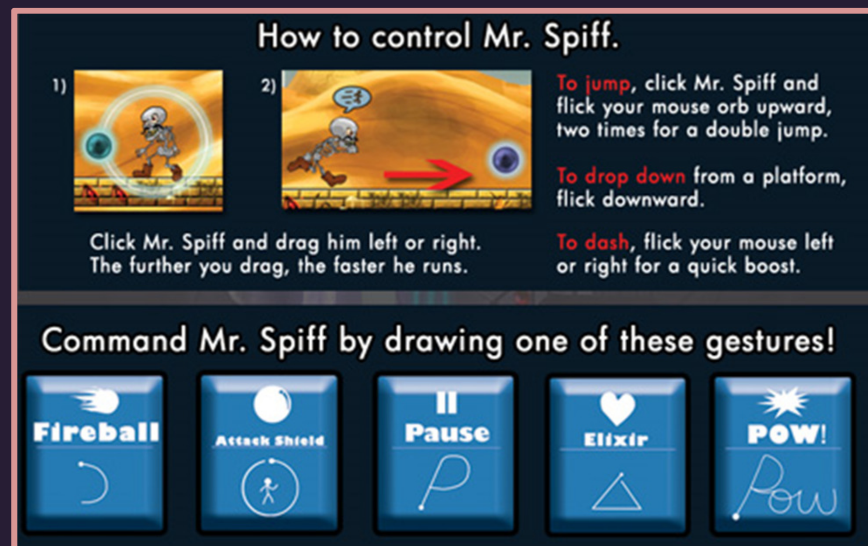
1. MOTIVATION

Pen and finger gesture input

My Word Coach (DS)



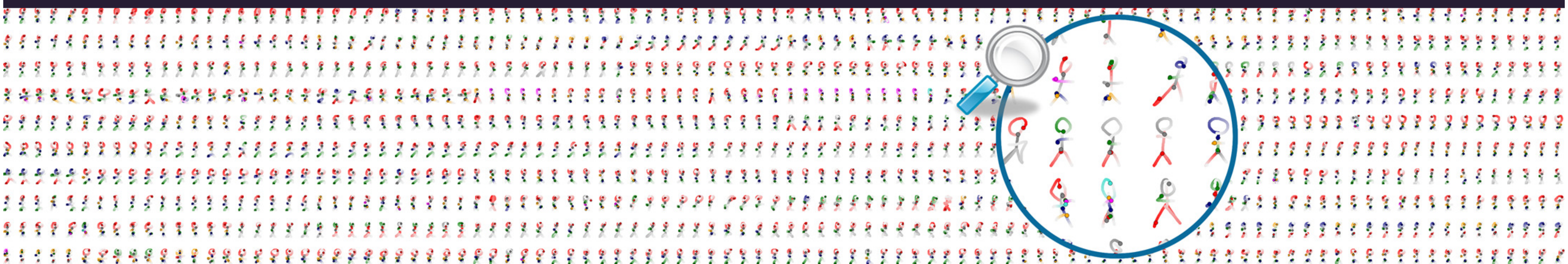
Obenkyo (Android)



Mr. Spiff's Revenge (PC)

Understanding how users make gestures

Understanding how users make gestures



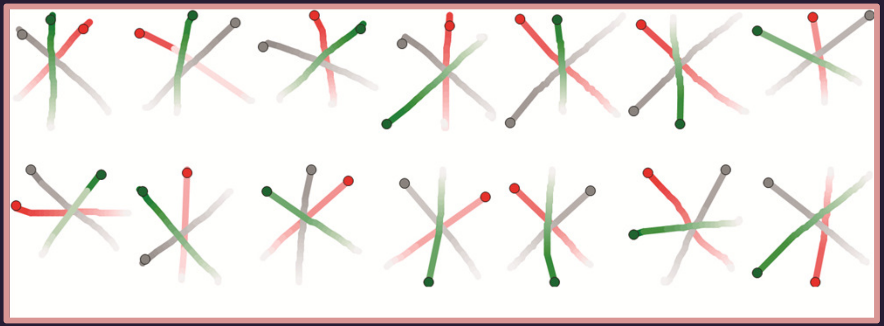
Nearly 1,000 ways that 34 different users made a 'person' gesture [NicIcon].

Goals:

- Understanding how users actually make gestures
- Knowing how consistently users make gestures
- Designing and developing better interaction and recognition

Absolute accuracy measures

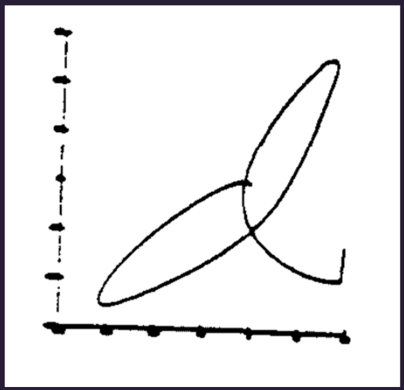
- Number of strokes
- Path length
- Gesture area (bounding box)
- Cosine of starting angle
- Cosine of ending angle
- Line similarity
- Global orientation
- Total turning angle
- Sharpness
- Curviness
- Production time
- Average speed



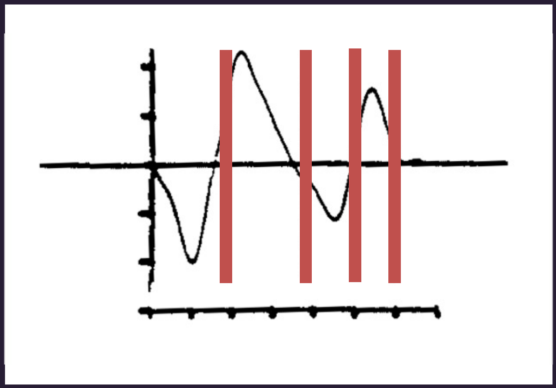
A 3-stroke asterisk can be made in up to 48 ways; here are the 14 drawn by 20 actual users over 200 samples [MMG].

Source: Anthony et al, GI 2013

Human motor control theory



A handwritten ‘d’.



Velocity profile of the ‘d’.

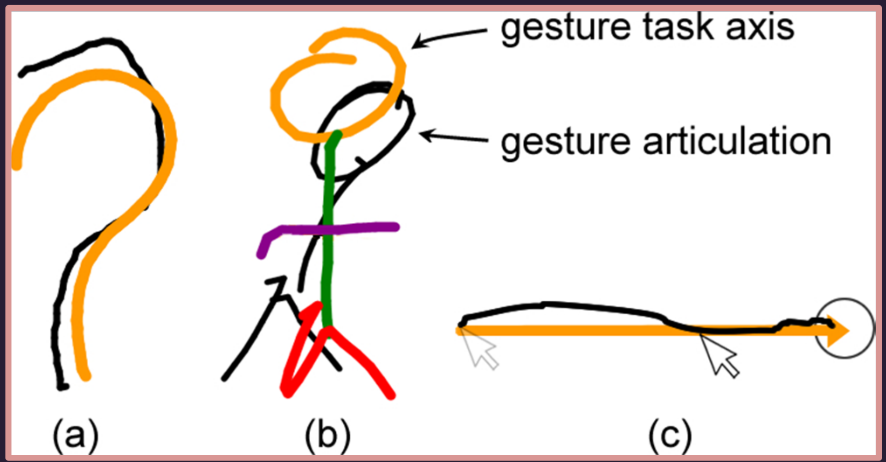
Source: http://www.google.com/patents/WO1994009447A1

2. THE RELATIVE ACCURACY MEASURES

Basis of the relative accuracy measures

Mouse **pointing path accuracy** measures of [Mackenzie et al, CHI 2001] inspired our work.

We derived **12 new measures** based on those that work for the nuances of a gesture path (not a straight line).

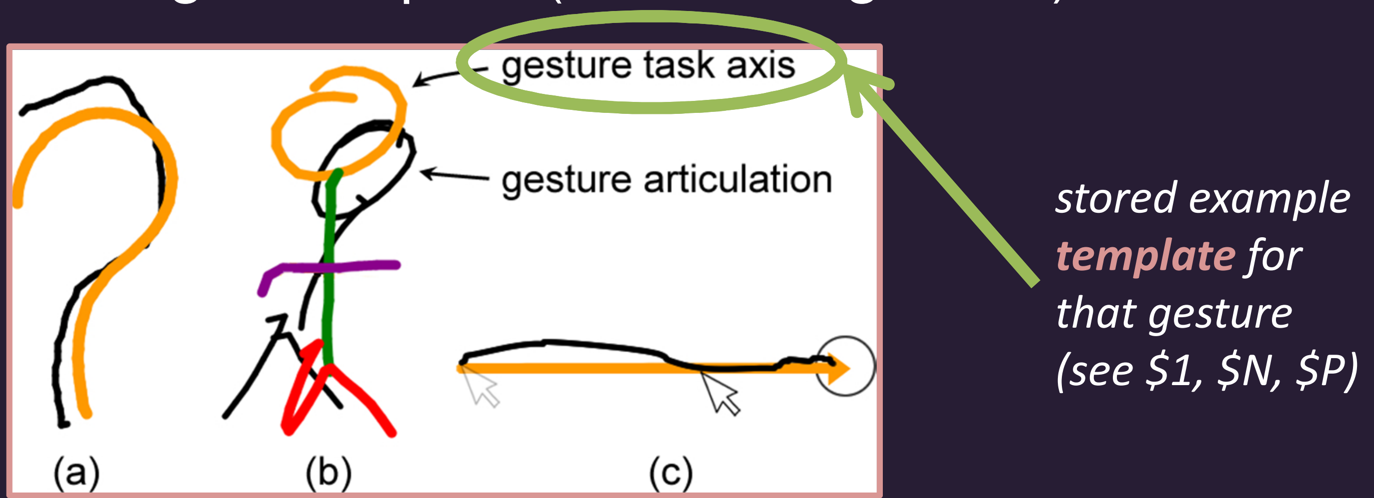


Examples of how we conceptualize the comparison of a gesture path to a stored template path (a and b), just as Mackenzie et al. examined mouse pointing paths compared to the optimal straight line path (c).

Basis of the relative accuracy measures

Mouse **pointing path accuracy** measures of [Mackenzie et al, CHI 2001] inspired our work.

We derived **12 new measures** based on those that work for the nuances of a gesture path (not a straight line).



Examples of how we conceptualize the comparison of a gesture path to a stored template path (a and b), just as Mackenzie et al. examined mouse pointing paths compared to the optimal straight line path (c).

Defining the relative accuracy measures

Three types of measures:

- **Geometric** – evaluate the deviation of the candidate gesture from the task axis in terms of shape distance, and capture tendencies of the users to stretch and bend strokes during articulation.
- **Kinematic** – evaluate articulation differences in the time domain, and capture how fluent or smooth the articulated path is in terms of production time and speed.
- **Articulation** – measures how consistent users are in producing the individual strokes of gestures.

Defining the relative accuracy measures

Our 12 measures focus on either "**Error**" or "**Variability**":

- **Error** is defined with respect to the stored template – how different is the user's path from the stored path.
- **Variability** is defined with respect to the standard deviation of the differences between the stored template and user's gesture – high values indicate lots of relative variability.

The Relative Accuracy Measures

Geometric accuracy:

- **Shape Error** – average absolute deviation of the candidate gesture points from the task axis in terms of Euclidean distance.
- **Shape Variability** – standard deviation of the distances between the points of the candidate and the task axis.
- **Length Error** – measures users' tendencies to "stretch" gesture strokes with respect to the gesture task axis.
- **Size Error** – measures users' tendencies to "stretch" gesture strokes in terms of the gesture area size (bounding box).
- **Bending Error** – measures users' tendencies to "bend" the strokes of the articulated gesture with respect to gesture task axis.
- **Bending Variability** – standard deviation of the differences in turning angle per gesture point in the sample and task axis.

The Relative Accuracy Measures

Kinematic accuracy:

- **Time Error** – measures the difference in articulation time (total duration) between the candidate and the gesture task axis.
- **Time Variability** – standard deviation of the differences between timestamps measured at each individual point on the gesture path.
- **Speed Error** – measures the difference in the speed profiles of the candidate and the gesture task axis.
- **Speed Variability** – standard deviation of the local differences between the speed profiles.

The Relative Accuracy Measures

Articulation accuracy:

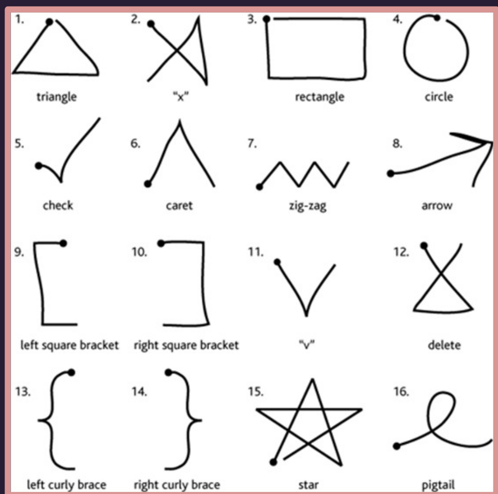
- **Stroke Count Error** – difference in the number of strokes between the candidate and the gesture task axis.
- **Stroke Ordering Error** – indicator of stroke ordering accuracy, computed as the absolute difference between the $1 cost measure and the $P cost measure.

3. RESULTS

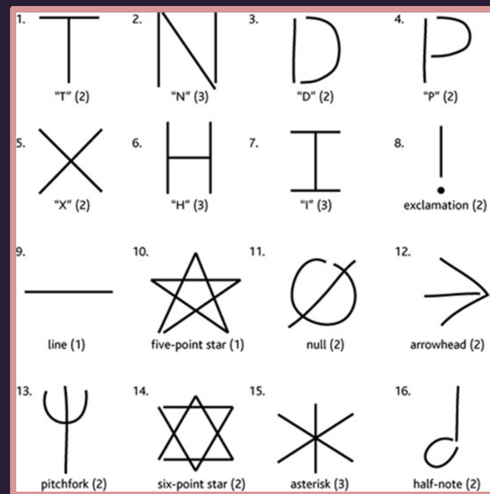
Validating the relative accuracy measures

Research question: what new understanding of users' gesture articulation patterns, and their relationships to recognition performance, can our relative accuracy measures reveal?

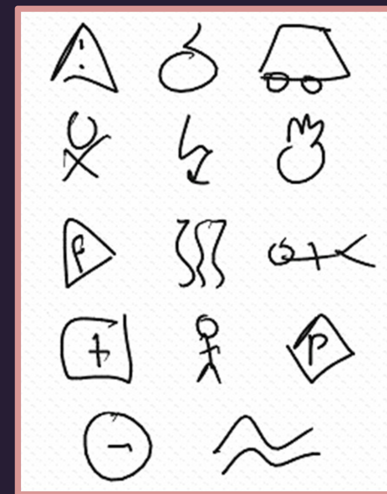
Datasets used



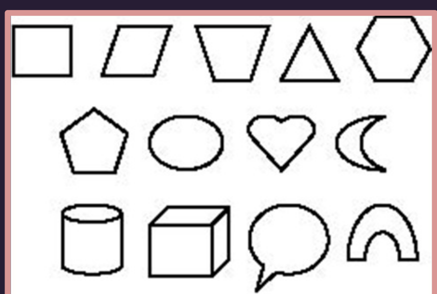
$1 dataset



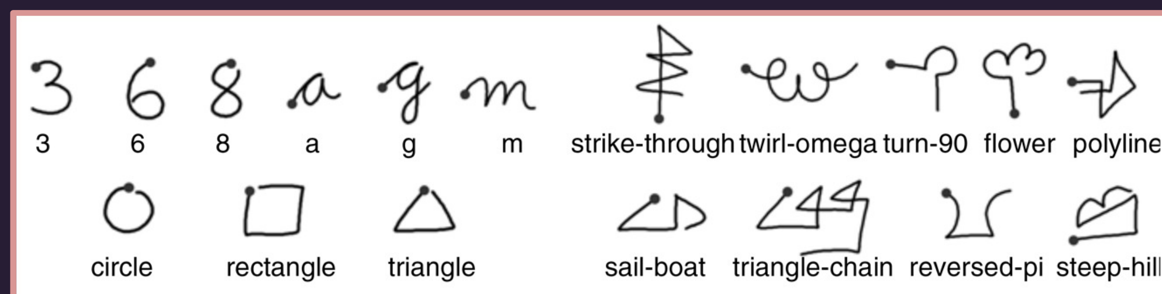
MMG dataset



NicIcon



HHReco



Vatavu et al, INTERACT 2011 simple and complex gestures

(Download links for all datasets are included in the paper.)

Universitatea Ştefan cel Mare Suceava

Computing the gesture task axis

All 12 relative accuracy measures are computed relative to the **gesture task axis**.

We explored three ways to compute the task axis and compare them in the paper to pick the best way.

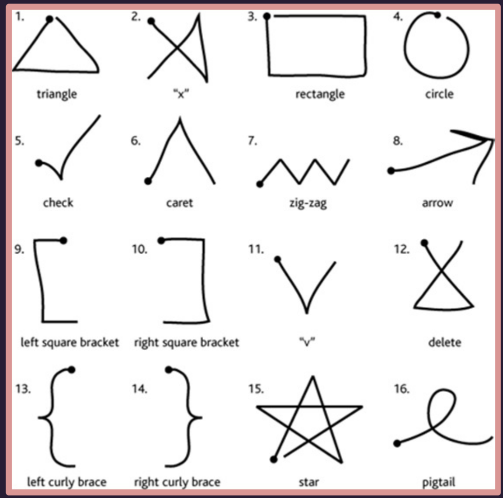
The gesture task axis for each shape (shown in orange) is one example articulation of the gesture from the stored set of templates (thin lines).

Example case study: $1 dataset

$1 dataset [Wobbrock et al, UIST 2007]

- 10 participants, 10 samples, 16 gestures
- 3 speeds ('fast', 'medium', 'slow')

Original study found effect of articulation speed on recognition accuracy: **medium was best accuracy**.



$1 dataset

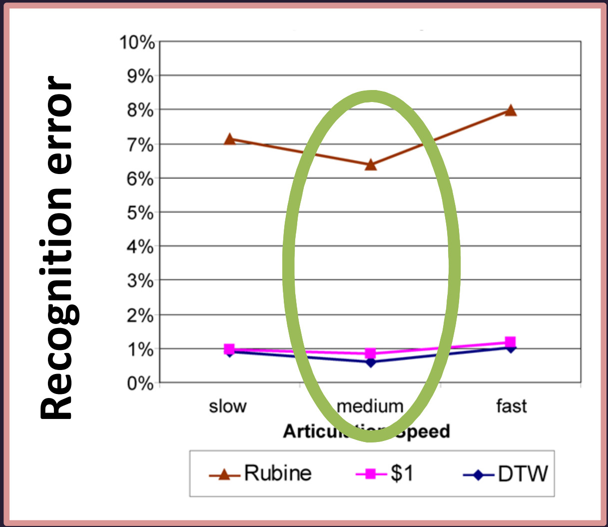
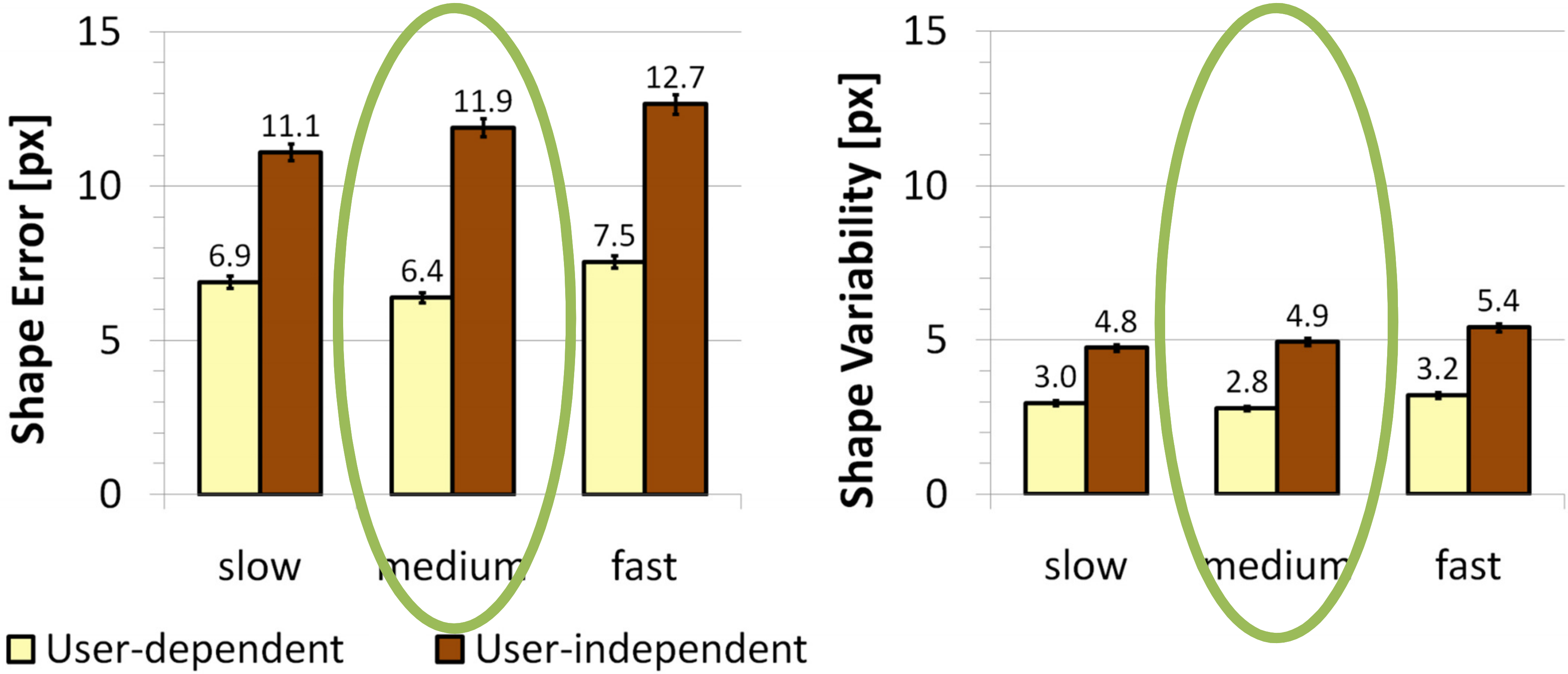


Figure 9(b) from Wobbrock et al, UIST 2007: "Recognition error rates as a function of articulation speeds (lower is better)."

Case study: $1 dataset – digging deeper

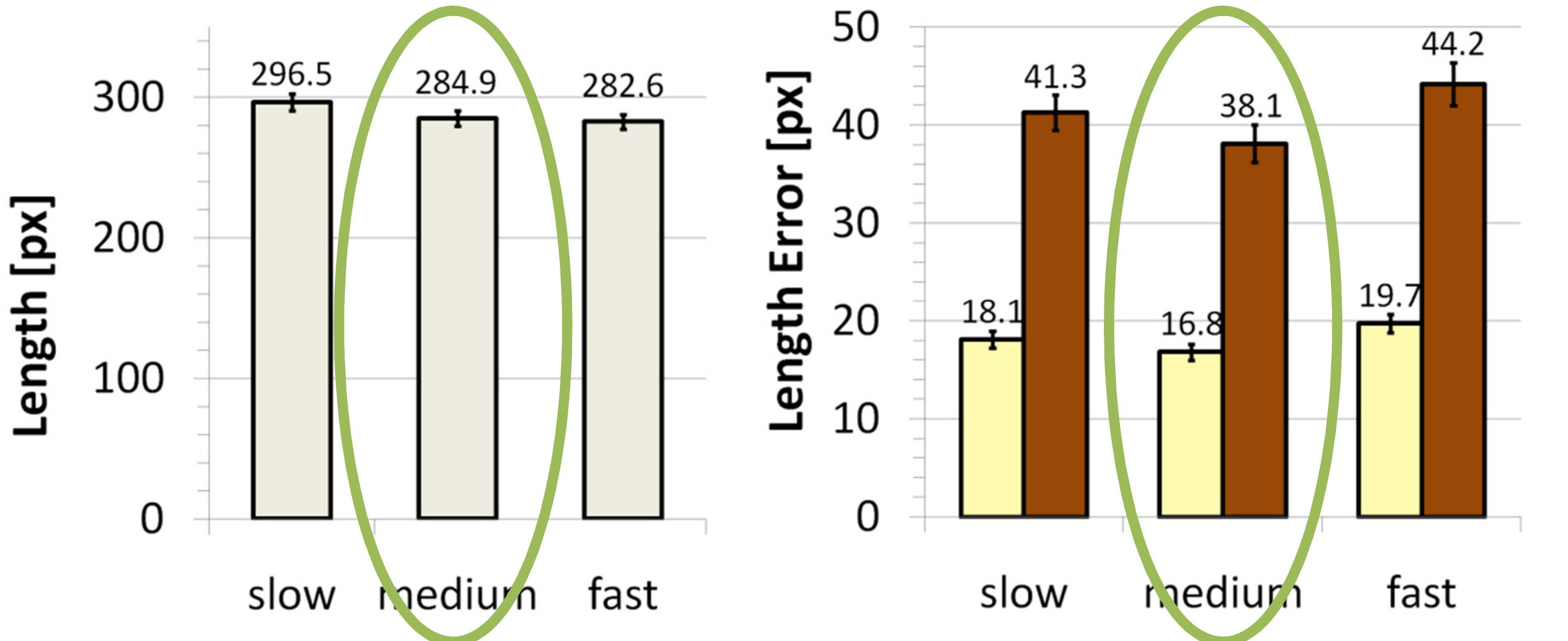


Shape Error and Variability reflect differences in recognition rate (e.g., medium gestures have the smallest errors and are the most accurately recognized in the user-dependent case).

Universitatea Ştefan cel Mare Suceava

Vatavu, Anthony, & Wobbrock. Relative Accuracy Measures for Stroke Gestures, ICMI 2013

Case study: $1 dataset – new discoveries



Absolute gesture length (left) compared to relative error measures (right). Note how the relative measures reveal articulation characteristics not captured by absolute measures (e.g., even though fast gestures are shorter in path length, their lengths varied the most).

Case study: $1 dataset – implications

1. Understanding how users make gestures with **different articulation speeds** (e.g., situational, user characteristics, etc.)
2. Use **recognizers that are robust** to the specific relative measure change we see in our datasets
3. **Design gesture sets** to avoid gestures more significantly affected by users' tendencies to stretch their strokes

4. CONCLUSION

Extensions to this work

Comparing different **user populations**…

Comparing different **touchscreen devices**…

Use by industry **UI/UX professionals**…

- Recent tweets about this work show possible pick-up already!

Nick Finck, Senior UX Manager, Amazon

- 5 Retweets and 19 Favorites by UX/UI professionals around the world…

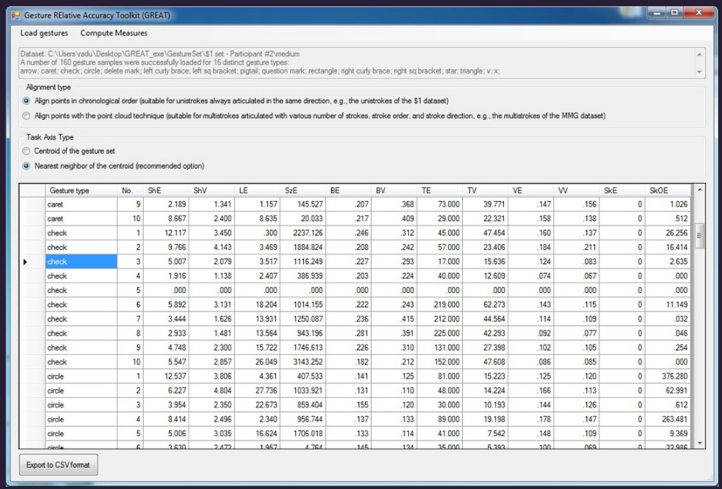
GREAT tool available for download!

GREAT = **Gesture RElative Accuracy Toolkit**

DLL that implements the relative accuracy measures

- **Open source and customizable** for those wishing to investigate further and even develop new measures

Demo application showing how to use the DLL to compute the measures for your own datasets:



GREAT screenshot

Summary

- ✓ We've been continuing our efforts to **understand how users make gestures**.
- ✓ We introduce **12 relative accuracy measures** that reveal nuances in differences between gestures.

- ✓ These new measures can support existing findings about how users make gestures and **how this impacts recognition accuracy**.
- ✓ These new measures can go beyond what is possible with absolute features to **uncover more fine-grained understandings** of the impact of gesture variance.

Questions?

Contact:

- Radu-Daniel Vatavu, PhD
  - vatavu@eed.usv.ro
- Lisa Anthony, PhD
  - lanthony@cise.ufl.edu
- Jacob O. Wobbrock, PhD
  - wobbrock@uw.edu



Online demo / download of GREAT tool:

- http://depts.washington.edu/aimgroup/proj/dollar/great.html

References

1. Anthony, L., and Wobbrock, J. O. A lightweight multistroke recognizer for user interface prototypes. GI ’10, Canadian Inf. Proc. Soc. (Toronto, Ont., Canada, 2010), 245–252.
2. Anthony, L., Vatavu, R.-D., and Wobbrock, J. O. Understanding the consistency of users’ pen and finger stroke gesture articulation. Canadian Inf. Proc. Soc. (Toronto, Ont., Canada, 2013), 87–94.
3. Hse, H., and Newton, A. Recognition and beautification of multi-stroke symbols in digital ink. Computers & Graphics 29, 4 (2005), 533–546.
4. MacKenzie, I. S., Kauppinen, T., and Silfverberg, M. Accuracy measures for evaluating computer pointing devices. CHI ’01, ACM (New York, NY, USA, 2001), 9–16.
5. Vatavu, R.-D., Vogel, D., Casiez, G., and Grisoni, L. Estimating the perceived difficulty of pen gestures. INTERACT’11, Springer-Verlag (Berlin, Heidelberg, 2011), 89–106.
6. Willems, D., Niels, R., van Gerven, M., and Vuurpijl, L. Iconic and multi-stroke gesture recognition. Patt. Rec. 42, 12 (2009), 3303–3312.
7. Wobbrock, J. O., Wilson, A. D., and Li, Y. Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes. UIST ’07, ACM (New York, NY, USA, 2007), 159–168.