Singular Value Decomposition is a Valid Predictor of Stroke Importance in Reading Chinese

Hsueh-Cheng (Nick) Wang[1], Bernhard Angele[2], Elizabeth Schotter[2], Jinmian Yang[2], Dan Simovici[1], Marc Pomplun[1], and Keith Rayner[2]
[1] Department of Computer Science, University of Massachusetts at Boston, USA
[2] Department of Psychology, University of California, San Diego, USA

Introduction

Alphabetic languages

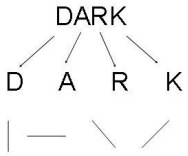
- Certain letter positions are more important for identification than others.
- Beginning letters are very important for word identification and reading comprehension.
  - -Replacing (Rayner and Kaiser,1975), visually degrading (Jordan, Thomas, Patching, & Scott-Brown, 2003) and transposing letters (Rayner, White, Johnson & Liversedge, 2006) is most disruptive in external (particularly initial) positions.

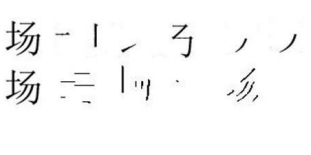
Chinese

- Not all strokes are of equivalent importance and **configuration is important**.
- Removing strokes hinders word identification Tseng, Chang and Wang (1965)
  - -Deleting strokes that retain character configuration is least disruptive, deleting final strokes was more disruptive, deleting initial strokes was most disruptive (Yan et al., 2011).

Representation of words

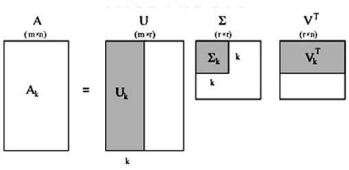
- Alphabetic words can be broken down into components (e.g., letters, features; McClelland & Rumelhart, 1981).
- Strokes are dots, lines, or curves. Some are combinations of these.





Singular Value Decomposition (SVD)

- A mathematical method in linear algebra which "orders" components of a matrix in terms of informativeness.



a 他每天早晨都到操场上锻炼身体
b 他每天早晨都到操场上锻炼身体
c 他每天早晨都到操场上锻炼身体
d 他每天早晨都到操场上锻炼身体
e 他每天早晨都到操场上锻炼身体

Methods

1 他每天早晨都到操场上锻炼身体
2 他每天早晨都到操场上锻炼身体
3 他每天早晨都到操场上锻炼身体
4 他每天早晨都到操场上锻炼身体

- Twelve native speakers of Chinese at UMass Boston participated.
- SR Research EyeLink Remote system with a sampling of 1000 Hz
- Subjects read each sentence from Yan et al. (2011) in 4 conditions:
  1 All retained
  2 70% most important segments retained
  3 70% least important segments retained
  4 70% random segments retained
- 48 sentences in random order. Each sentence was followed by a true or false question.

- When most informative segments were retained reading was no different from reading with all segments retained.
- When least informative segments were retained reading was most disrupted.
- When random segments were retained reading was moderately disruptive.
- Data suggest that most important segments in characters are not necessarily the initial segments and can be identified using SVD.