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

Hsueh-Cheng Wang

Department of Computer Science, University of Massachusetts at Boston, USA

Bernhard Angele, Elizabeth Schotter, Jinmian Yang

Department of Psychology, University of California, San Diego, USA

Dan Simovici, Marc Pomplun

Department of Computer Science, University of Massachusetts at Boston, USA

Keith Rayner

Department of Psychology, University of California, San Diego, USA

Chinese characters are written in a specific, non-arbitrary stroke order, which may reflect a stroke's importance. Previous research on stroke removal showed that removing the initial strokes from characters made them harder to read than removing the final strokes or the shortest strokes [Yan et al., 2011, Reading and Writing, 1-29]. Singular value decomposition (SVD; a method of decomposing the matrix of pixels representing a character in order to degrade it, but retain the most important segments) may be a better way to estimate which elements of a character are most important for identification. In the present study, characters were decomposed into segments (i.e., vertical, horizontal, and diagonal lines) and the importance of each segment's contribution to character configuration was determined by SVD. Subjects read each sentence from the Yan et al. study either with all or 70% of the segments retained. The retained segments were the most important, the least important, or randomly selected. When the most important segments were retained, subjects read as fast as when all were retained. When the least important segments were retained, reading fluency was reduced. The results suggest that SVD is a psychologically valid way to capture Chinese character configuration.