Fixation detection in noisy eye-movement data

Hessels, Roy S; Niehorster, Diederick C; Kemner, Chantal; Hooge, Ignace T C

2016

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Fixation detection in noisy eye-movement data
Roy S. Hessels¹, Diederick C. Niehorster², Chantal Kemner¹, & Ignace T.C. Hooge¹
¹ Utrecht University, ² Lund University

Eye-tracking research in infants and older children has gained a lot of momentum over the last decades. While eye-tracking research in these participants groups has become easier with the advance of the remote eye tracker, this often comes at a cost of poorer data quality compared to research with well-trained adults (Hessels, Andersson, Hooge, Nyström, & Kemner, 2015; Wass, Forssman, & Leppänen, 2014). Current solutions for automatic detection of fixations in eye-tracking data, both those provided by eye-tracker manufacturers and the research community, are not built for data from infants and older children. As a result, the output of such algorithms are often unreliable when the data are noisy. The purpose of the present work was to address the need for an algorithm capable of labeling fixations across a wide range of noise levels in data where periods of data loss may occur. Here we proposed and evaluated a new algorithm designed specifically for eye-movement data of varying data quality: Identification by 2-Mean Clustering (I2MC). In comparison with seven other state-of-the-art event-detection algorithms, we found that the number of fixations, mean fixation duration and standard deviation of fixation duration by the I2MC algorithm are most robust to increases in noise level and data loss. The I2MC algorithm is applicable to fixation labeling in situations where the data quality may be low, for instance when working with infants, school children, or certain patient groups. The I2MC algorithm may be used when the noise and data loss levels are markedly different between trials and/or subjects; the output should be comparable despite these differences in noise and data loss levels.