EEG 및 시선추적 신호 기반 가독성 평가 상관분석

윤세휘 [1], 송명주 [1], 황세미 [1, 2†]
[1] 모두의연구소, [2] 연세대학교 원주연세의료원

soyshoong@gmail.com, myungjoo.song12@gmail.com, semi.hwang11@gmail.com

Correlation Analysis of Readability Assessment Using EEG and Eye-Tracking Signals

Sehwi Yoon[1], MyungJoo Song[1], Semi Hwang[1, 2]
[1]MODULABS, [2] YONSEI UNIVERSITY WONJU HEALTH SYSTEM

Abstract

Conventional readability indices rely on surface-level linguistic features and often fail to capture the cognitive effort involved in reading. We present a cognitively grounded framework that integrates linguistic, eye-tracking, and EEG signals for readability estimation. Using the ZuCo 2.0 dataset, we combined sentence length, fixation duration, and oscillatory power in the alpha and theta bands with human readability ratings. Correlation analyses show that multimodal features outperform unimodal measures, with the integration of linguistic and cognitive signals yielding stronger alignment with human judgments. These results suggest that readability emerges from the interaction between text and cognition with practical implications for adaptive learning and accessibility.

Ⅰ. INTRODUCTION

Readability, defined as the ease with which a reader can process and understand text, is important in education, accessibility, and NLP. Conventional readability metrics such as FRES, FKGL, and SMOG rely on surface-level linguistic features like sentence length and word difficulty. While simple and interpretable, they may not adequately capture the cognitive demands readers experience in natural settings.

Psycholinguistic and cognitive neuroscience research shows that reading reflects an interaction between text and cognition. Eye-tracking measures (fixation count, first fixation duration, gaze duration) and EEG oscillations in the theta (4– 7 Hz) and alpha (8– 12 Hz) bands are sensitive indicators of cognitive load. These findings suggest that cognitive signals can enrich traditional readability indices.

However, few studies systematically combine linguistic, eye-tracking, and EEG features to model human readability judgments. Prior work offers partial evidence, but integration across modalities remains limited. To address this gap, we propose a multimodal approach to readability estimation, hypothesizing that combining linguistic and cognitive signals will align more closely with human judgments than unimodal metrics.

Ⅱ. METHODS

This study employed the ZuCo 2.0 dataset, which originally contains 738 English Wikipedia sentences with simultaneous EEG and eye-tracking recordings from 18 participants. For our analysis, we selected a subset of 150 sentences. Since the dataset does not include readability annotations, we additionally collected human readability ratings. A total of 50 raters participated, with each rater evaluating 15 sentences, and each sentence receiving ratings from three independent raters. The averaged score per sentence was used as the ground truth for analysis.

Feature extraction was performed across three modalities. Linguistic features included sentence length, the number of difficult words, and classical readability indices such as FRES and FKGL. Eye-tracking features consisted of the number of fixations (nFix), first fixation duration (FFD), and total gaze duration (GD), which reflect attentional allocation during reading.

EEG features were derived from the alpha and theta frequency bands. Specifically, we extracted the mean, maximum, and standard deviation of band power across frontal (Fz, F3, F4) and parietal (Pz, P3, P4) channels. Variables denoted in Table 1 correspond to the maximum and standard deviation of oscillatory power, respectively.

All features were normalized at the participant level and aggregated by sentence. Normalization was performed using z-score transformation at the participant level to reduce inter-individual variability and to ensure comparability across subjects. To evaluate their relevance, Pearson correlation analysis was conducted between the extracted features and human readability ratings, both individually and in multimodal combinations.

	|coefficient|
Flesch Reading Ease	0.296
Flesch-Kincaid Grade	0.50
sentence length	**0.610**
# difficult words	0.527
nFix	0.173
FFD	0.226
GD	0.226
α_{std}	0.18
α_{max}	0.191
θ_{std}	0.191
θ_{max}	0.194
EEG_{std}	0.168
sentence length + FFD	0.611
sentence length + α_{std}	0.612
sentence length + α_{max}	0.618
sentence length + θ_{std}	0.612
sentence length + θ_{max}	0.618
sentence length + α_{max} + θ_{max}	0.6240
sentence length + FFD + α_{max} + θ_{max}	**0.6242**

Table 1. Pearson correlation coefficients between human readability ratings and various linguistic, eye-tracking, and EEG-based features

Ⅲ. RESULTS AND DISCUSSIONS

Table 1 summarizes the correlation coefficients between human readability ratings and the extracted features. Among linguistic features, sentence length showed the highest correlation (r = 0.610), followed by the number of difficult words (r = 0.527) and FKGL (r = 0.50). Eye-tracking signals demonstrated moderate but consistent associations, as both first fixation duration and gaze duration achieved correlations of r = 0.226, while the number of fixations was slightly lower at r = 0.173. EEG-derived features exhibited smaller yet meaningful contributions, with theta maximum reaching r = 0.194, aligning with prior evidence that oscillatory power in the theta and alpha bands reflects working memory and processing load.

Notably, multimodal combinations yielded higher alignment with human judgments. As shown in Table 1, the integration of sentence length, first fixation duration, and alpha/theta maxima produced the highest observed correlation (r = 0.6242). While this represents only a moderate association, it nevertheless outperforms unimodal feature sets.

produced the strongest correlation (r = 0.6242), outperforming any unimodal feature set. These multimodal gains are consistent with Hollenstein et al. (2019), who demonstrated that EEG and eye-tracking features contribute complementary information to linguistic features in comprehension-related tasks. Compared to François & Miltsakaki (2012), who reported moderate improvements over traditional readability formulas using NLP-based features, our results suggest that direct incorporation of cognitive signals offers an even stronger alignment with human judgments.

The observed theta-band contributions align with prior findings that theta oscillations reflect working memory load during sentence processing. This suggests that readability judgments are sensitive not only to text difficulty but also to cognitive effort required for semantic integration.

Overall, these findings suggest that readability is not solely determined by textual properties but emerges through the interaction of text with human cognitive processes. By incorporating neural and behavioral signals, the proposed approach provides a more human-centered and cognitively grounded method of readability assessment, with potential applications in adaptive learning, accessibility, and inclusive language technologies.

Ⅳ. CONCLUSION

This study presented a cognitively grounded analysis of readability assessment by integrating linguistic features with EEG and eye-tracking signals. The findings demonstrate that multimodal features, reflecting neural load and attentional effort, align more closely with human judgments than traditional surface-level metrics. By highlighting the interaction between text and cognition, this work advances the understanding of readability assessment and provides initial empirical evidence, though future work should employ regression or machine learning models to validate predictive power.

ACKNOWLEDGMENT

This research was supported by Brian Impact Foundation, a non-profit organization dedicated to the advancement of science and technology for all.

References

[1] R. Flesch, "A new readability yardstick," Journal of Applied Psychology, vol. 32, no. 3, pp. 221-233, 1948.

[2] J. P. Kincaid, R. P. Fishburne, R. L. Rogers, and B. S. Chissom, "Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy enlisted personnel," Research Branch Report 8-75, Naval Technical Training Command, Millington, TN, 1975.

[3] G. H. McLaughlin, "SMOG grading—A new readability formula," Journal of Reading, vol. 12, no. 8, pp. 639-646, 1969.

[4] J. S. Chall and E. Dale, Readability Revisited: The New Dale- Chall Readability Formula. Brookline, MA: Brookline Books, 1995.

[5] T. François and E. Miltsakaki, "Do NLP and machine learning improve traditional readability formulas?," Proc. 1st Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR), pp. 49-57, 2012.

[6] N. Hollenstein, M. Tröndle, C. Zhang, and N. Langer, "ZuCo 2.0: A dataset of physiological recordings during natural reading and annotation," Proc. Int. Conf. on Language Resources and Evaluation (LREC), pp. 138-146, 2020.

[7] N. Hollenstein, M. Tröndle, M. Plomecka, S. Kiegeland, Y. Özyurt, L. A. Jäger, and N. Langer, "The ZuCo benchmark on cross-subject reading task classification with EEG and eye-tracking data," Frontiers in Psychology, vol. 13, 1028824, 2023.

[8] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis," Brain Research Reviews, vol. 29, no. 2-3, pp. 169-195, 1999.

[9] K. G. Klein, S. Frenkel, O. Shubi, and Y. Berzak, "Eye tracking based cognitive evaluation of automatic readability assessment measures," arXiv preprint arXiv:2502.11150, 2025.