Koh, E., et al. (Eds.) (2026). Proceedings of the 2nd International Conference on Learning Evidence and Analytics. Asia-Pacific Society for Computers in Education.

ENGFORTHAI+: Mitigating Morphological Omission with Static Visual Scaffolding and Adaptive Content Generation

Narabodee RODJANANANT[a], Phurinat POLASA[b] & Nattapol KRITSUTHIKUL[c*]
[a]*Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Thailand*
[b]*Department of Computer Engineering, Faculty of Engineering, King Mongkut's University of Technology Thonburi, Thailand*
[c]*National Electronics and Computer Technology Center (NECTEC), Thailand*
*nattapol.kritsuthikul@gmail.com

Abstract: While recent advancements in L1-tuned Automatic Speech Recognition (ASR) have significantly improved the diagnosis of segmental pronunciation errors, L2 learners continue to exhibit persistent morphological deficits. Specifically, Thai EFL learners frequently omit inflectional markers due to the high cognitive load of real-time morpho-syntactic planning. This study introduces "Static Visual Scaffolding" (SVS), a novel intervention designed to bridge the gap between declarative knowledge and procedural speech production using dynamic, AI-generated passage sentences. Unlike traditional Intelligent Tutoring Systems that provide reactive, post-turn feedback, the SVS framework acts as a "Heads-Up Display" (HUD) for the learner. As the learner reads contextually unique AI-generated scripts, the system utilizes low-latency streaming ASR to anticipate and mitigate grammatical spoken errors. The HUD projects subtle visual cues (e.g., a spectral highlight on specific morphemes) before the learner articulates the target verb, effectively preempting habitual errors. We evaluated this approach through a randomized controlled trial involving 30 Thai university students engaging in the oral reconstruction of AI-generated passages. Results indicate that the SVS group achieved a statistically significant reduction ($\eta_p^2 = 0.41$) in grammatical spoken errors events compared to a control group receiving standard post-hoc corrections. Furthermore, analysis suggests that combining static cueing with infinite, AI-driven content reduces extraneous cognitive load, allowing learners to maintain fluency while attending to form. These findings propose a paradigm shift in Computer-Assisted Language Learning (CALL): moving from error detection to anticipatory cognitive offloading during complex speech tasks.

Keywords: Mispronunciation Detection and Diagnosis (MDD), Computer-Assisted Language Learning (CALL), Reflective Learning

1. Introduction

Effective L2 communication relies not only on segmental articulation but also on the precise execution of morpho-syntactic features. For Thai EFL learners, omitting inflectional markers (e.g., plural /-s/, past tense /-ed/) is a pervasive issue, often attributed to the high cognitive load of processing real-time speech rather than a lack of declarative knowledge. This difficulty is compounded by L1 interference, as Thai is an isolating language that lacks the inflectional suffixation found in English (Pongpairoj, 2002).

Recent advancements in Computer-Assisted Pronunciation Training (CAPT) have significantly improved the diagnosis of these errors. Rodjananant et al. (2025) demonstrated that L1-tuned Automatic Speech Recognition (ASR) enhances sentence-level error detection for Thai learners by accounting for specific accent variations.

Despite these advances, a critical limitation persists as current systems operate on a reactive paradigm, providing feedback only after an error has occurred. While useful for post-hoc reflection, this approach fails to scaffold the learner during the critical moments of articulatory planning (Bořil et al., 2024). The limitations of reactive systems are best understood through Cognitive Load Theory (Sweller, 1988), which suggests that effective interventions should provide personalized, adaptive support that scaffolds learning without overwhelming working memory (Afzaal et al., 2024; Lehman et al., 2020). This is particularly crucial for L2 learners, who exhibit significant variability in morphological prediction compared to native speakers due to working memory constraints (Lozano-Argüelles et al., 2023). This convergence of pedagogical need and technological capability creates an opportunity to design systems that are proactive rather than reactive.

To address this gap, this study proposes the ENGFORTHAI+ framework, equipped with "Static Visual Scaffolding" (SVS). This approach shifts the focus from error detection to anticipatory cognitive offloading. By visually priming target forms within adaptive, AI-generated passages, the system aims to reduce extraneous cognitive load, allowing learners to maintain fluency while attending grammatical accuracy.

2. System Architecture

The ENGFORTHAI+ framework uses cyclical architecture that supports iterative improvement in pronunciation and morphological accuracy. As shown in Figure 1, the system forms a closed loop in which learner speech is analyzed by a Thai-tuned ASR model, scored for target grammatical omissions, and used to generate the next set of practice materials.

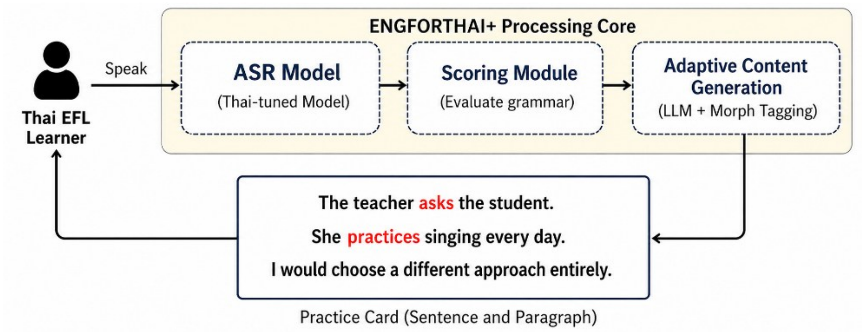

Figure 1. *System architecture of the ENGFORTHAI+ framework.*

The system begins with a diagnostic phase using a Fixed Diagnostic Paragraph (FDP) to establish baseline morphological competence. This static passage contains a high density of target inflectional markers, including:

1. Third-Person Singular verbs (e.g., walks, chooses, washes).
2. Regular Past Tense markers (e.g., walked, played, decided).
3. Plural Nouns in complex clusters (e.g., tests, desks).

As FDP is identical for all users, it provides a standardized stress test of omission patterns and phonological difficulty. The resulting error profile serves as seed data for the practice loop, where the system generates personalized sentences and short passages targeting the learner's weak grammatical forms alongside known phonological challenges for Thai learners. These materials are presented as practice cards with *Static Visual Scaffolding*, which highlights target morphemes to increase their salience during oral reconstruction.

Following the initial FDP evaluation, the system initiates an adaptive practice loop where Google's Gemini 3 Flash generates CEFR-appropriate texts that combine the learner's specific grammatical deficits with common Thai phonological challenges. These exercises scale from isolated sentences to full paragraphs to test morphological robustness under increasing cognitive load, utilizing static visual scaffolding (red text) to draw immediate attention to target morphemes. The learner's oral reconstruction is then processed by a Thai-tuned ASR and scoring module, which calculates grammatical error rates; if omissions exceed a predefined threshold, the system flags these specific features and feeds them back into the LLM prompt to generate highly personalized content for subsequent practice cycles.

3. Experiment

To evaluate the efficacy of ENGFORTHAI+ in reducing morphological omissions, we conducted a classroom-based randomized controlled trial with 30 Thai undergraduates equally distributed across CEFR levels A2, B1, and B2 (n = 10 per level). The study examined the effect of "Static Visual Scaffolding" (SVS) on oral production accuracy. All participants provided informed consent before taking part in the study. Participants were then assigned to two groups using stratified randomization by CEFR level:

- Control Group (Visual Cues Disabled): Participants read practice passages presented in standard black text.
- Experimental Group (Visual Cues Enabled): Participants read the same content, but with the SVS system (target morphemes highlighted in red and underlined) active.

The procedure had three phases: (1) a pre-test using the FDP without visual scaffolding or feedback to establish baseline Grammatical Omission Rate (GOR), (2) four 20-minute ENGFORTHAI+ practice sessions, and (3) a post-test using a new FDP comparable to the pre-test. Both experiment groups used the adaptive sentence generation system so that the experiment isolated the effect of visual cues. Subsequently, audio from the pre-test and post-test was processed by the Thai-tuned ASR model and verified by human annotators.

4. Results

The metric for evaluation was the Grammatical Omission Rate (GOR), defined as the percentage of target inflectional markers omitted during speech. Lower scores indicate higher accuracy. Table 1 presents the descriptive statistics and results of the one-way Analysis of Covariance (ANCOVA) conducted to determine the effect of the SVS system.

The pre-test scores were significantly related to the post-test scores, $F(1, 27) = 35.12$, $p < 0.001$, $\eta_p^2 = 0.57$. After adjusting for these pre-existing differences, there was a statistically significant effect of the intervention on morphological accuracy, $F(1, 27) = 18.81$, $p < 0.001$. The effect size was large ($\eta_p^2 = 0.41$), indicating that the visual scaffolding intervention accounted for approximately 41% of the variance in the final omission rates, separate from the learners' initial aptitude. This suggests that the static visual cues successfully offloaded the cognitive burden of morphological monitoring, allowing learners to produce the target forms accurately.

Table 1. *Descriptive statistics and ANCOVA results for Grammatical Omission Rate (GOR).*

Group	N	Pre-test M (SD)	Post-test M (SD)	Adjusted Post-test M
Control	15	0.425 (0.048)	0.382 (0.043)	0.380
Experimental	15	0.418 (0.049)	0.342 (0.020)	0.344

Further analysis was conducted to determine how learners of different proficiency levels responded to the intervention. Basic learners (CEFR A2) exhibited the highest baseline omission rate (~55%). However, they also showed the largest magnitude of improvement when using the system, with error rates dropping to ~18%. This indicates that lower-proficiency learners, who often lack the cognitive resources to monitor form, benefit most from external visual scaffolding. Intermediate learners (CEFR B2) started with a lower baseline omission rate (~25%). Although they still improved with the system (dropping to ~8%), their percentage gain was lower than the A2 group.

Despite the modest sample size, the randomized design, covariate-adjusted analysis, and large effect size lend support to the robustness of the observed effects within this learning analytics driven adaptive framework.

5. Discussion

The ASR system served not only as an evaluative tool but also as a diagnostic instrument that revealed systematic phonological patterns. Omission errors were not random; the model frequently flagged verbs ending in complex final consonant clusters, especially voiceless stops followed by sibilants (e.g., asks /æsks/, stops /stɒps/, texts /teksts/). This finding aligns with Thai phonotactic constraints, which do not permit final consonant clusters, and suggests that learners may know the grammatical rule but still struggle with the articulatory demand of producing these forms. The result supports the role of the Adaptive Content Generation module in targeting high-friction words rather than generic vocabulary.

Post-experiment interviews further suggested that the effectiveness of SVS can be interpreted through the Noticing Hypothesis. Participants in the Experimental Group reported that the static visual cues interrupted their habitual reading flow and made target morphemes more salient. As one student explained, “I usually guess the word and move on, but the red color forced me to stop and prepare my mouth to say the s"'. This indicates that the visual cue helped bridge the gap between declarative knowledge and procedural speech production by prompting learners to consciously attend to forms they would otherwise omit.

6. Conclusion, Limitation & Future Work

In this study, we enhanced a personalized learning platform to address the persistent challenge of morphological omission among Thai EFL learners, specifically focusing on the production of past tense markers (Rodjananant et al., 2025). By integrating a real-time ASR pipeline with Static Visual Scaffolding (SVS), the ENGFORTHAI+ system aimed to bridge the gap between declarative grammar knowledge and procedural speech production.

Our results show that ENGFORTHAI+ significantly reduces morphological omission rates by using SVS to support procedural speech production. While the current study was conducted with a specific cohort of Thai university students, the significant effect sizes suggest that SVS remains a promising intervention for supporting proceduralization. A key remaining limitation is that SVS cannot resolve the underlying articulatory deficit associated with producing final consonant clusters that are absent from learners’ L1 phonology. Future work will therefore integrate an interactive articulatory model that provides real-time visual feedback on tongue and lip positioning. We also plan to investigate whether visual scaffolding can be gradually reduced across practice sessions while maintaining learner performance. In addition, we plan to conduct a longitudinal study with more diverse participant groups to determine whether the benefits of visual scaffolding led to durable gains in morphological accuracy.

References

Afzaal, M., Zia, A., Nouri, J., & Fors, U. (2024). Informative Feedback and Explainable AI-Based Recommendations to Support Students’ Self-regulation. *Technology, Knowledge and Learning*, *29*(1), 331–354. https://doi.org/10.1007/s10758-023-09650-0

Bořil, T., Kocjančič Antolík, T., & Hofmann, A. (2024). Acoustic and Articulatory Visual Feedback in Classroom L2 Vowel Remediation. System, 120, 103204.

Lehman, B., Gu, L., Zhao, J., Tsuprun, E., Kurzum, C., Schiano, M., Liu, Y., & Tanner Jackson, G. (2020). Use of Adaptive Feedback in an App for English Language Spontaneous Speech. *Artificial Intelligence in Education*, *12163*, 309–320.

Lozano-Argüelles, C., Sagarra, N., & Casillas, J. V. (2023). Interpreting experience and working memory effects on L1 and L2 morphological prediction. *Frontiers in Language Sciences*, *1*. https://doi.org/10.3389/flang.2022.1065014

Pongpairoj, N. (2002). The acquisition of the English plural morpheme by L1 Thai learners: A case of the Failed Functional Features Hypothesis. Chulalongkorn University.

Rodjananant, N., Polasa, P., Pompech, P. N., Saengchan, T., Boonma, K., & Kritsuthikul, N. (2025). Improve English Pronunciation at Sentence Level for Thai EFL Learners With Thai Automatic Speech Recognition Model. *International Conference on Learning Evidence and Analytics*.

Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, *12*(2), 257–285.