Asian Journal of Technology and Applied Sciences
ISSN: 2301-xxxx (online) | IF: 10.01 | Vol: 1, issue 1
Available: http://asiansciencejournal.org

INTEGRATING SELF-MONITORING METACOGNITIVE TOOLS WITH VOICE AI FOR REDUCED SPEECH HESITATION IN EFL CLASSROOMS

Ibodullayeva Diyora Orifjon qizi
Samarqand davlat chet tillar instituti talabasi
diyoraibodullayeva50@gmail.com

Abstract: Speech hesitation hinders spontaneous oral communication among EFL learners. While Conversational Voice AI provides interactive practice, standalone use lacks mechanisms to foster self-regulated speech production. This study evaluated the efficacy of integrating self-monitoring metacognitive tools (visual pause analytics, reflection logs, and goal sheets) with Voice AI to reduce speech hesitation among intermediate EFL university students. Utilizing an 8-week mixed-methods design, participants were divided into experimental (Voice AI + metacognitive tools) and control (Voice AI alone) groups. Pre- and post-intervention monologues were analyzed acoustically via Praat software for filled pauses, unfilled long pauses, and speech rate, alongside qualitative thematic analysis of reflection logs. Results demonstrated a significant reduction in filled and unfilled pauses and a substantial increase in speech rate in the experimental group compared to the control group. Qualitative findings confirmed that visual feedback heightened temporal self-awareness, shifting learners from silent freezing to active paraphrasing strategies. Integrating metacognitive tools with Voice AI transforms passive technology use into active self-regulation, offering a scalable framework for enhancing EFL oral fluency.

Keywords: Voice AI, Metacognition, Speech Hesitation, EFL Oral Fluency, Self-Regulated Learning, Acoustic Analysis.

Introduction. Speaking fluency remains one of the most challenging competencies for English as a Foreign Language (EFL) learners to attain. According to Levelt's (1989) foundational speech production model, oral communication demands real-time cognitive processing, requiring simultaneous conceptualization, formulation, and articulation. In EFL environments where natural immersion is limited, learners frequently experience speech hesitation – manifested through unfilled pauses, filled pauses (e.g., uh, um), and false starts (Kormos, 2006). While brief pauses are a natural component of cognitive planning in native speech (Trofimovich & Baker, 2006), excessive hesitation reflects high cognitive load, communication anxiety, and inefficient self-monitoring mechanisms (Gilquin, 2008). In recent years, rapid advancements in Artificial Intelligence (AI) have introduced Conversational Voice AI tools (such as automated speaking tutors, speech recognition systems, and LLM-powered voice assistants) into Computer-Assisted Language Learning (CALL) (Godwin-Jones, 2021; Kohnke, 2023). These applications provide learners with scalable, low-anxiety environments for interactive speaking practice. However, pedagogical observations indicate that interactive AI tools alone do not automatically guarantee fluency development (Jeon, 2022). Without explicit structural guidance, learners often engage in passive interactions with Voice AI, prioritizing task completion over the active refinement and regulation of their oral production. Metacognition – the awareness and regulation of one's own learning processes – plays a decisive role in autonomous language acquisition (Flavell, 1979; Wenden, 1998). Grounded in Flavell's metacognitive framework and Sweller's (1988) Cognitive Load Theory, self-monitoring mechanisms allow learners to assess their speech in real-time, identify sources of hesitation (such as vocabulary retrieval gaps versus grammatical uncertainty), and deploy compensatory strategies (Oxford, 2017). Integrating explicit self-

monitoring tools – such as visual pause analytics, post-task reflection logs, and structured goal-setting rubrics – into Voice AI environments bridges the gap between interactive practice and active, self-regulated speech reflection (Zhang & Goh, 2020). Although substantial research has investigated the isolated effects of Voice AI on pronunciation and speaking motivation (Hwang & Chang, 2021), as well as the benefits of metacognitive instruction in writing and reading comprehension, limited empirical research has examined the synergistic effect of combining real-time self-monitoring metacognitive tools with Voice AI explicitly to reduce speech hesitation in EFL contexts. It remains unclear whether self-monitoring feedback prompts can transform AI-driven speaking sessions into self-regulated learning opportunities that demonstrably improve speech fluency and reduce temporal hesitation markers. This study aims to bridge this gap by examining how integrating self-monitoring metacognitive tools with Voice AI impacts the reduction of speech hesitation markers among intermediate EFL learners over an 8-week period. Specifically, the study seeks to answer the following research questions: RQ1: Does the integration of self-monitoring metacognitive tools with Voice AI significantly reduce speech hesitation (filled and unfilled pauses) in intermediate EFL learners compared to using Voice AI alone? RQ2: How does the combined use of Voice AI and metacognitive monitoring tools affect learners' overall speech rate and self-correction performance? RQ3: What are the learners' perceptions regarding the role of self-monitoring tools in managing their cognitive load and speaking anxiety during AI-mediated speech tasks?

Methodology. This study utilized a mixed-methods research design (Creswell & Creswell, 2018), combining quantitative and qualitative data, to examine whether integrating self-monitoring metacognitive tools with Voice AI helps reduce speech hesitation among EFL learners. A total of 60 intermediate-level (CEFR B1) university EFL students participated in this 8-week investigation. All participants were screened using the Oxford Quick Placement Test to ensure a homogeneous initial language proficiency. The participants were randomly divided into two equal groups: an experimental group (n = 30) practicing speaking using Conversational Voice AI combined with self-monitoring metacognitive tools, and a control group (n = 30) practicing speaking using Conversational Voice AI alone. Both groups conducted their practice using a Conversational Voice AI application powered by large language models as an interactive speaking partner (Kohnke, 2023). The AI engaged students in informal conversations and short discussions on everyday academic topics suitable for B1 level learners. In addition to the Voice AI sessions, the experimental group utilized three specific self-monitoring tools grounded in Oxford's (2017) strategic self-regulation framework: visual pause feedback displaying silent pauses longer than 2.0 seconds (Goldman-Eisler, 1968; Kormos, 2006), post-task reflection logs where students recorded the causes of their hesitations (such as vocabulary retrieval gaps, grammar uncertainty, or anxiety), and weekly goal sheets helping them set concrete targets for smoother delivery.

The empirical intervention lasted 8 weeks, consisting of two 45-minute practice sessions per week. During Week 1, all 60 participants completed a pre-test by recording a 2-minute spontaneous monologue on a prompt regarding daily technology. From Weeks 2 to 7, both groups completed their assigned speaking practice tasks, with the experimental group additionally reviewing their visual pause feedback and submitting reflection logs after each session. In Week 8, participants completed a post-test recording under identical conditions on a comparable prompt, and 12 selected students from the experimental group took part in semi-structured qualitative interviews.

Data analysis comprised both quantitative acoustic measurements and qualitative thematic processing. Pre- and post-test speech recordings were analyzed using Praat acoustic software

(Boersma & Weenink, 2023) alongside automated speech rate scripts (De Jong & Wempe, 2009). Three key temporal speech markers were measured: filled pauses (frequency of vocalized hesitations like uh and um per minute), unfilled long pauses (frequency of silent gaps exceeding 2.0 seconds per minute), and overall speech rate (syllables spoken per minute). Statistical procedures were conducted in IBM SPSS using paired-samples t-tests for within-group comparisons and ANCOVA to establish between-group differences (Plonsky & Oswald, 2014). Finally, qualitative reflection logs and interview transcripts were analyzed using Braun and Clarke's (2006) thematic analysis to capture learner perspectives on cognitive load management and self-regulation.

Results. The analysis of the speech recordings and qualitative data yielded positive evidence regarding the efficacy of combining self-monitoring metacognitive tools with Voice AI. Prior to the intervention, both groups exhibited similar baseline rates of speech hesitation markers during the pre-test. The experimental group demonstrated a mean of 8.42 filled pauses per minute, 5.12 unfilled pauses longer than 2.0 seconds per minute, and an average speech rate of 92.4 syllables per minute (SPM). Similarly, the control group recorded a mean of 8.58 filled pauses per minute, 5.28 unfilled long pauses per minute, and a speech rate of 90.8 SPM. These baseline measurements confirmed that both groups started from equivalent levels of oral fluency prior to the intervention. Following the 8-week intervention, quantitative acoustic analysis via Praat revealed a statistically significant reduction in speech hesitation markers for the experimental condition compared to the control condition. The experimental group experienced a substantial decrease in filled pauses (e.g., uh, um), dropping to 3.15 per minute, which represented a statistically significant within-group improvement ($t = 13.48$, \; $p < .001$, \; $d = 3.48$). The control group also demonstrated a decrease in filled pauses to 6.10 per minute ($t = 5.62$, \; $p < .01$, \; $d = 1.46$), though the magnitude of change was noticeably lower. An ANCOVA controlling for baseline pre-test scores confirmed a significant main effect between conditions ($F(1, 57) = 48.32$, \; $p < .001$, \; \eta_p^2 = .459). Furthermore, unfilled silent pauses exceeding 2.0 seconds dropped significantly in the experimental group to 1.82 per minute ($t = 12.15$, \; $p < .001$, \; $d = 3.08$), whereas the control group retained an average of 4.25 long silent pauses per minute. Consequently, the experimental group's overall speech rate increased by 34.8% to 124.6 SPM ($p < .001$), compared to 102.1 SPM in the control group. Qualitative findings derived from the 480 post-task reflection logs and 12 semi-structured interviews provided context for these quantitative improvements, revealing three main operational shifts in learners' cognitive processing. First, participants noted that visual pause feedback heightened their temporal awareness, helping them identify previously unnoticed silent gaps during lexical retrieval (Zhang & Goh, 2020). Second, reflection logs demonstrated a cognitive strategy shift; rather than freezing or using silent gaps when encountering vocabulary difficulties, learners actively deployed compensatory strategies such as filler phrases (e.g., "What I mean is...") and circumlocution (Oxford, 2017). Finally, participants expressed that interacting with Voice AI in a low-stakes environment combined with personalized goal setting reduced communication anxiety and encouraged active self-regulation during spontaneous speech production.

Discussion. The primary objective of this study was to investigate whether integrating self-monitoring metacognitive tools with Voice AI effectively reduces speech hesitation among EFL learners. The overall findings demonstrate that while Voice AI alone provides a beneficial platform for speaking practice, pairing it with structured metacognitive scaffolding yields significantly greater improvements in oral fluency and cognitive self-regulation. The quantitative results clearly indicate a substantial reduction in both filled (uh, um) and long unfilled pauses (> 2.0s) in the experimental group compared to the control group. Enhanced

Cognitive Processing: The dramatic reduction in filled pauses (from 8.4 to 3.1 per minute) suggests that visual pause feedback and reflection logs helped learners streamline their speech planning mechanisms. Reduction of Silent Gaps: The decrease in unfilled long pauses (down to 1.8 per minute) indicates that self-monitoring enabled students to bridge cognitive processing gaps more efficiently. Increased Speech Rate: The 34.8% increase in overall speech rate (up to 124.6 SPM) aligns with Kormos (2006) and Segalowitz (2010), demonstrating that reducing hesitation latency directly translates to enhanced temporal speech fluidity.

The observed outcomes support key theoretical frameworks in second language acquisition (SLA) and educational psychology:

Cognitive Load Theory (Sweller, 1988): Speaking in an L2 places high demands on working memory. Voice AI offers a low-anxiety environment, which lowers the affective filter (Krashen, 1982). Simultaneously, metacognitive self-monitoring tools break down speech delivery into manageable components, reducing cognitive overload during real-time speech production.

Self-Regulated Learning (Oxford, 2017; Flavell, 1979): Without metacognitive prompts, learners tend to use Voice AI passively, focusing only on message completion. Introducing post-task reflection logs transformed passive practice into active self-regulation, prompting learners to identify the root causes of their hesitations (e.g., lexical gaps vs. grammatical uncertainty). Strategic Compensation over "Freezing": Visual pause feedback acted as a mirror for learners' temporal speech habits (Zhang & Goh, 2020). As qualitative reflections revealed, learners shifted from complete silence or vocalized pause markers (uh/um) toward functional compensatory strategies, such as paraphrasing and filler expressions ("What I mean is..."). The findings offer practical guidance for educators and instructional technology designers: Avoiding Passive AI Usage: Language instructors should avoid using Conversational AI as a standalone "plug-and-play" tool. AI practice must be structured with explicit goal-setting and reflection mechanisms. Visualizing Performance Data: Providing learners with visual acoustic feedback (such as pause graphs) bridges the gap between how fluent learners feel they are and how fluent they actually sound.

Scaffolding Compensatory Strategies: Teachers should explicitly teach filler phrases and rephrasing techniques so students can replace silent hesitation with strategic communication tools. Despite positive findings, several limitations should be acknowledged:

Intervention Duration: The 8-week timeframe was sufficient to observe immediate temporal gains, but long-term retention of these fluency habits requires longitudinal follow-up research.

Sample Specificity: Participants were limited to intermediate (B1) university students. Results may vary for beginner (A2) or advanced (C1) EFL learners. Task Type Constraint: Data collection relied primarily on individual monologue tasks; future studies should evaluate peer-to-peer interactive tasks.

Conclusion. This study evaluated the impact of integrating self-monitoring metacognitive tools with Conversational Voice AI to reduce speech hesitation and enhance oral fluency among intermediate EFL learners. Over an 8-week intervention, the combination of interactive AI practice with visual pause feedback, reflection logs, and structured goal-setting led to significant reductions in both filled (uh, um) and long unfilled pauses, alongside a notable increase in overall speech rate. The findings highlight that while Conversational Voice AI offers an accessible, low-anxiety environment for speaking practice, technology alone is insufficient for optimal fluency development. Adding metacognitive self-monitoring tools transforms passive AI interaction into an active, self-regulated learning process.

Learners shift from silent freezing during vocabulary retrieval toward strategic paraphrasing and fluent communication.

In conclusion, integrating metacognitive scaffolding into AI-assisted language learning provides an effective, scalable pedagogical framework. Future research should explore the long-term retention of these fluency gains across various proficiency levels and examine how these strategies transfer to real-world, peer-to-peer communication settings.

REFERENCES

1. Boersma, P., & Weenink, D. (2023). Praat: doing phonetics by computer [Computer program]. Version 6.3.10. http://www.praat.org/
2. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa
3. Creswell, J. W., & Creswell, J. D. (2018). Research design: qualitative, quantitative, and mixed methods approaches (5th ed.). SAGE Publications.
4. De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and calculate speech rate automatically. Behavior Research Methods, 41(2), 385–390. https://doi.org/10.3758/BRM.41.2.385
5. Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second language fluency: Judgments on different tasks. Language Learning, 54(4), 655–679. https://doi.org/10.1111/j.1467-9922.2004.00285.x
6. Flavell, J. H. (1979). Metacognitive development. American Psychologist, 34(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906
7. Gilquin, G. (2008). Hesitations in native and non-native speech: A corpus-based approach. Language and Computers, 63(1), 122–148.
8. Godwin-Jones, R. (2021). Emerging technologies: Tools and trends in generative AI for language learning. Language Learning & Technology, 25(3), 4–19.
9. Goldman-Eisler, F. (1968). Psycholinguistics: Experiments in spontaneous speech. Academic Press.
10. Hwang, W. Y., & Chang, H. R. (2021). Facilitating EFL speaking performance and learning retention using mobile voice agents. Computers & Education, 162, 104085. https://doi.org/10.1016/j.compedu.2020.104085
11. Jeon, J. (2022). Exploring the potential of AI chatbots for language learning: What we know and what we should investigate. Journal of Educational Computing Research, 60(7), 1695–1717. https://doi.org/10.1177/07356331221082213
12. Kohnke, L. (2023). L2 learners' interactions with artificial intelligence conversational agents: A scoping review. Journal of Computer Assisted Learning, 39(3), 748–761. https://doi.org/10.1111/jcal.12767
13. Kormos, J. (2006). Speech production and second language acquisition. Lawrence Erlbaum Associates.
14. Krashen, S. (1982). Principles and practice in second language acquisition. Pergamon Press.
15. Levelt, W. J. M. (1989). Speaking: From intention to articulation. MIT Press.
16. Oxford, R. L. (2017). Teaching and researching language learning strategies: Self-regulation in context (2nd ed.). Routledge.
17. Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. Language Learning, 64(4), 878–912. https://doi.org/10.1111/lang.12079