Proceedings of ICEIPI 2026 Symposium: Inclusive Education, Cultural Transformation, and the Ethical Dimensions of Learning
DOI: 10.54254/2753-7048/2026.LD32652

A Study on the Impact of AI-Enabled Multimodal Teaching Methods on Oral English Expression of Junior High School Students

Enyang Guo

School of Foreign Languages and Literature, Suzhou University of Science and Technology, Suzhou, China
13812799249@163.com

Abstract. In the context of globalization, oral English proficiency has become increasingly important, and the junior high school stage is a crucial period for cultivating students' oral English skills. However, traditional English oral teaching lacks real-life contexts, monotonous teaching methods, and insufficient personalized guidance. Artificial intelligence has injected new vitality into multimodal teaching and provided new solutions to these problems. This study adopts a literature review method, systematically retrieved mainstream databases such as CNKI, Science Citation Index, and Google Scholar, and reviewed relevant literature on artificial intelligence, multimodal teaching, English oral teaching, and junior high school English published in the past decade. It analyzed the theoretical basis, the current status of application, the impact on students' oral English, and the existing challenges. This study shows that multimodal teaching based on artificial intelligence is based on constructivism, situated cognition, and input-output hypotheses. It creates a virtual, immersive environment and provides personalized support and intelligent feedback, thereby effectively improving students' oral skills and their willingness to communicate. However, there are still challenges, such as technological dependence, the digital divide, changes in teacher roles, and ethical issues. Future research should optimize human-machine collaboration, track long-term effects, and improve the assessment system.

Keywords: artificial intelligence (AI), multimodal teaching, junior high school English, oral expression, literature review

1. Introduction

In the context of globalization, oral English communication skills have become an important part of junior high school students' core competencies. However, its teaching still faces common problems, such as a lack of context, reliance on single teaching methods, insufficient personalized guidance, and imperfect evaluation mechanisms [1]. With the rapid development of AI, integrating it with multimodal teaching becomes an important solution.

This study uses a literature review to systematically analyze the theoretical basis, current status, effects, and future trends of AI-enabled multimodal teaching in junior high school oral English,

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2753-7048/2026.LD32652

aiming to provide references for related research and practice.

It is theoretically significant for enriching the IT-supported teaching framework and practically valuable for teachers and educational decision-makers.

2. Theoretical foundation of AI-enabled multimodal oral language teaching

2.1. Multimodal discourse analysis and teaching theory

The multimodal discourse analysis theory posits that meaning is jointly constructed by various symbolic resources such as language, images, and sounds, and communication relies on the collaborative efforts of multiple senses [2]. In this regard, the underlying principle behind the use of AI in teaching oral communication, especially in the context of junior high school English language learning, is that oral communication is a type of process that involves multiple perceptual inputs. By utilizing AI technology, teachers can incorporate multiple types of resources (such as voice, pictures, and situations) into the process of teaching oral communication, which is consistent with the fundamental conception of "collaborative communication in multiple ways using different senses" found in this theory. It helps junior high school students understand the meaning of language and improve the authenticity and fluency of their oral expression, laying a theoretical foundation for further exploration of its teaching impact.

2.2. Constructivism and situated cognition theory

Constructivism and Situated Cognition Theory suggest that language learning is not a passive reception process, but rather one in which learners actively construct meaning in real, interactive situations [3]. In other words, it is about "learning by doing". The multi-modal teaching enabled by AI technology today is precisely based on this concept.

Changes in teaching methods, such as incorporating VR or dynamic scenes into teaching processes (for example, creating immersive experiences in 3D worlds rather than just through reading plus static textbook content); expanding upon previous classes so that students do not feel they have to do it all alone; reducing fears associated with making mistakes, being laughed at for having done so and having the fear that their anxiety may impact their performance while participating in the learning process. Moreover, those AI intelligent agents with "personality" - they have fixed personalities, speak naturally, and can provide immediate feedback. When you chat with them, it feels just like interacting with a real person, with tone, emotion, and a genuine sense of communication. As a result, students no longer mechanically repeat and imitate; they actually dare to speak and express themselves. Through repeated human-machine interactions, language knowledge gradually becomes their own ability. This system is handy for junior high school students, as it can help them improve the fluency and accuracy of their oral language and, more importantly, make them more willing to speak English.

2.3. Input hypothesis and output hypothesis

Krashen's input hypothesis states that comprehensible input is the core condition for language acquisition [4]. That is, you need to come into contact with materials slightly above your current level but still understandable and comprehensible. Only in this way can your language skills truly improve. If the materials are too easy, there will be no progress; if they are too complicated, you will be discouraged directly.

DOI: 10.54254/2753-7048/2026.LD32652

The AI-powered multimodal teaching precisely addresses this issue. The system dynamically evaluates your present location from your oral language proficiency, vocabulary size, and grammar comprehension using information obtained above to properly select the correct learning tools to provide to all high school students working on learning oral language at their own pace through their own cognitive process and to prevent them from using the wrong materials to create inefficiencies while learning. Swain's output hypothesis further emphasizes that language output and feedback play a crucial role in acquisition. Learners can test their language assumptions during active output and identify problems and correct deviations through feedback [5]. In essence, it is about you speaking out. Only when you make a mistake, and someone points it out, can you check whether your understanding is correct, then make adjustments and improvements.

The AI-powered multimodal platform is also competent in this regard. It can provide real-time scores for your spoken language, offering instant feedback across multiple dimensions, including pronunciation, intonation, grammar, and word choice. Moreover, it will not put you under the same pressure as a teacher would. This non-judgmental feedback helps you focus on the issue itself, prompting you to continually reflect, correct, and restructure your expression as you speak.

When these two theories are combined, they precisely form the underlying logic of AI-assisted oral language teaching: the input should be appropriate, and the output should be effective. This is indeed very helpful for improving the accuracy, fluency, and practical communication skills of junior high school students.

3. The application model and practice of AI-powered multimodal oral language teaching

3.1. Technical support for teaching resources and environment creation

In the junior high school English oral communication class, AI can actually solve many longstanding problems.

Previously, the resources for oral communication classes were too limited, and the scenarios were not realistic enough. Students would repeatedly practice by merely reading texts and reciting dialogues, which was very dull. But now, with the use of AI, the situation is quite different.

The system can automatically select appropriate materials based on each student's oral language proficiency, the level they have reached, and their interests—text, images, audio, and animations [6]. These can be integrated to form a dynamic, personalized audio picture book library. Instead of just reading text, students can now combine audio and visuals, creating a rich mix of pictures and text. Learning becomes less boring, and students are more willing to participate and practice speaking actively. The foundation for this practice naturally gets laid.

What is more impressive is that AI can also generate some virtual characters with distinct personalities, such as "Marco Polo" [7], to allow students to engage in time travel, cross-cultural communication, or simulate e-commerce live broadcasts and daily conversations --- scenarios that are close to real life and future work. Students interact and communicate in a realistic context, and their oral communication skills, accuracy, and practicality can all be improved.

Overall, AI-powered multimodal teaching makes junior high school oral communication classes more interesting and effective. The practical value is quite obvious.

3.2. Oral practice with AI agents

At the level of innovative teaching processes and methods, AI technology has injected new vitality into multimodal teaching.

DOI: 10.54254/2753-7048/2026.LD32652

Using AI agents to practice oral English is actually one of the core strategies of AI-enabled multimodal teaching in junior high school English classes. Its advantages are obvious: it provides each student with personalized conversation scenarios to practice repeatedly, with less pressure and anxiety.

The main approach is to design an "avatar" for the AI—a unique personality, a rich background story, and natural-sounding speech. These all align well with the cognitive characteristics of junior high school students, making it more suitable for learning.

Research has shown that AI applications such as Talkpal.AI provide the ability to automatically change the level of difficulty in conversations based on individual students' spoken ability. They enhance students' fluency and pronunciation using three modes of engagement (audio, verbal and visual) [8]. More importantly, this model addresses the major pain points of traditional oral English classes: the inability to find real human coaches and the lack of practice opportunities [9].

The AI agent can provide immediate feedback and ask targeted questions, like a "little bit knowledgeable partner". During teacher-student interaction and human-machine conversations, it builds a scaffold for students, guiding them to standardize their expressions and enhance their ability to apply language in practice.

Overall, this provides a reliable, practical path for oral language instruction in junior high school English.

3.3. Real-time error correction and feedback of AI voice interaction technology

The real-time error correction function of AI voice interaction has indeed changed the old routine of junior high school English oral practice. Previously, teachers were too busy to provide timely feedback, or the feedback was too general, and students couldn't know exactly where they went wrong. Now, with AI, this problem can be largely solved.

Speech practice methodologies have been revolutionized due to real-time error identification/feedback of A.I. voice interaction technology [10]. Intelligent voice recognition technology enables systems/functions to instantly detect error(s) in pronunciation, grammar and fluency. Then, based on dimensions such as pronunciation standardization, sentence structure norms, and expression fluency, it will provide specific and actionable improvement suggestions. Whether you have a strong or weak foundation, it can match your level and meet your personalized needs.

The most crucial aspect is that this feedback is immediate. Once the student finishes speaking, they can immediately identify the problem and correct it promptly [11]. This immediate and detailed feedback enables students to precisely correct their mistakes through repeated attempts, thereby significantly improving the accuracy and fluency of their language expression. This is the tangible support that technology provides for improving oral skills.

4. The impact of AI-enabled multimodal teaching on junior high school students' English oral expression

4.1. The impact on oral language skills

In terms of its impact on oral language skills, AI technology demonstrates multi-dimensional facilitating effects.

Firstly, it can correct pronunciation and grammatical errors, standardize vocabulary usage, and make the expression more accurate. Secondly, the practice environment created by AI has less

DOI: 10.54254/2753-7048/2026.LD32652

pressure and allows for repeated practice. Students do not have to worry about making mistakes and being laughed at. Once they start speaking, their expressions become more fluent and coherent.

More importantly, AI can also help students enhance their pragmatic skills and deepen their understanding of the content. It uses multimodal technology to create a range of realistic cross-cultural scenarios that simulate communication in different situations [12]. It enables students to understand the cultural rules, social etiquette, and contextual differences behind the language, and learn to speak appropriately depending on the occasion. In the immersive interaction, students can gradually master the appropriateness of expression.

4.2. The impact on learning emotions and attitudes

Based on existing research, AI multimodal teaching does indeed make junior high school students more interested in speaking English. The abundant multimedia resources, fun AI interactions, and immediate feedback all work together to effectively stimulate learning motivation, making students genuinely want to learn.

Especially in enhancing willingness to communicate and boosting self-confidence, the "non-judgmental" environment created by AI is crucial [13]. In traditional classrooms, if you make a mistake, you fear being laughed at by your classmates, which can be very pressure. But when practicing with AI, there are no such concerns. You don't have to worry about losing face, and you can keep making mistakes and speaking boldly. Moreover, AI will immediately correct your mistakes and encourage you. Gradually, students won't be afraid to express themselves; they will be willing to communicate in English, and their willingness to communicate and self-confidence will naturally increase. This also lays a solid foundation for them to participate in oral activities in the future actively.

4.3. The impact on the ability of autonomous learning

In cultivating autonomous learning, the AI-enabled multimodal teaching method plays a significant role in advancing progress.

AI can customize an individualized learning path for each junior high school student based on their oral language foundation and progress [14]. Coupled with immediate feedback and precise data analysis, students can clearly understand where they are weak and how much they have improved, achieving effective self-monitoring.

Moreover, AI exercises are very flexible. Students can arrange their own time and adjust the intensity. Through repeated trial and error and correction, they gradually develop the habit of actively reflecting and adjusting their learning strategies, and their ability to independently plan and explore also improves accordingly.

5. Challenge and reflections

Although AI-powered multimodal oral language teaching has significant advantages, it also faces various challenges.

Among them, the technical challenges are the most prominent. First of all, while the conversations created by AI can seem very human like, they are not able to convey the same emotions that humans have [15]. The longer you talk to an AI, the more rigid and less 'human' the interaction becomes. In addition, the technology itself is not always reliable; issues such as voice recognition errors or system freezes do occur from time to time. When the practice is halfway

DOI: 10.54254/2753-7048/2026.LD32652

through and interrupted, it dramatically affects the experience. Moreover, the distribution of equipment is not balanced. Some students do not have the conditions at home and cannot use it at all.

Additionally, teachers' digital literacy is not keeping pace; students are overly dependent on technology, and there are ethical issues as well. These cannot be ignored. In summary, all these factors limit the large-scale promotion of AI oral language teaching.

6. Conclusion

This study conducted a systematic review of relevant literature and found that the multimodal teaching method based on artificial intelligence is based on constructivism, situated cognition, and the input-output hypothesis. Through virtual simulation environments and AI voice interaction technology, it enables "human-machine collaboration" for personalized learning. This model has been applied in various ways in junior high school English oral language teaching, significantly improving students' oral accuracy, expression fluency, and appropriateness of language use, while reducing language anxiety and enhancing learning interest and autonomous learning ability.

However, it still faces challenges such as technical stability, teachers' digital literacy, students' ability to filter information, and data ethics.

In the future, it is necessary to optimize the human-machine collaboration mechanism, improve the assessment system, and balance technical ethics and educational equity to promote the healthy and sustainable development of this model in junior high school English oral language teaching.

References

[1] Xu, Y. (2024). AIGC digital humans in improving junior high school students' oral English communication competence: An application study [Master's thesis, Southwest University]. China Doctoral Dissertations & Master's Theses Full-text Database. https: //doi.org/10.27684/d.cnki.gxndx.2024.004863\

[2] Zhang, D. (2009). Exploring a comprehensive theoretical framework for multimodal discourse analysis. Foreign Languages in China, 6(1), 24–30. https: //doi.org/10.13564/j.cnki.issn.1672-9382.2009.01.004

[3] Zhong, M. (2025). Building an AI intelligent agent oral practice model based on English textbook characters. Teaching and Management, (11), 55–58.

[4] Liu, Y. (2024). Research on multimodal oral English teaching in the context of human-computer dialogue examinations [Master's thesis, Southwest University]. China Doctoral Dissertations & Master's Theses Full-text Database. https: //doi.org/10.27684/d.cnki.gxndx.2024.004290

[5] Xie, Y. (2023). An action research on junior high school oral English teaching from the perspective of multimodal discourse analysis theory [Master's thesis, Hunan University of Science and Technology]. China Doctoral Dissertations & Master's Theses Full-text Database. https: //doi.org/10.27738/d.cnki.ghnkd.2023.000341

[6] Shi, J. (2024). Research on the application of English audio picture books in junior high school English listening and speaking teaching under multimodal theory [Master's thesis, Southwest University]. China Doctoral Dissertations & Master's Theses Full-text Database. https: //doi.org/10.27684/d.cnki.gxndx.2024.000517

[7] Chen, Y. (2025). Practical exploration of AI-empowered situational creation in secondary vocational English teaching: A case study of oral teaching for Unit 7 "Invention and Innovation" in Higher Education Press's secondary vocational English basic module (Book 2). Campus English, (49), 75–78.

[8] Hidayatullah, E. (2024). The impact of Talkpal. AI on English speaking proficiency: An academic inquiry. Journal of Insan Mulia Education, 2(1), 19-25.

[9] Zhang, Y., & Chen, Y. (2025). A comparative study of the effects of real-person interaction and real-person-AI interaction in business English oral communication from the perspective of sociocultural theory. Overseas English, (15), 99–103.

[10] Xu, H. (2025). Exploration of innovative models for junior high school oral English teaching based on "AI voice interaction." English Journal for Middle School Students, (44), 59–60.

[11] Jiang, J. (2024). Reflections on AI-empowered paths for oral English teaching. Journal of Inner Mongolia University of Finance and Economics, 22(6), 72–76. https: //doi.org/10.13895/j.cnki.jimufe.2024.06.010.

DOI: 10.54254/2753-7048/2026.LD32652

[12] Hai, Y. (2025). Analysis of students' language output competence in college oral English teaching from a multimodal perspective. Journal of Taiyuan Urban Vocational College, (9), 137–139. https://doi.org/10.16227/j.cnki.tycs.2025.0512
[13] Sun, Y., Liang, W., Wang, H., Li, P., Zhou, Y., & Gao, F. (2021). Practical research on using AI technology to improve students' oral expression ability in English teaching. In Proceedings of the 2021 Education Science Network Seminar (pp. 288–291). Shenzhen Longgang District Lucheng Foreign Language Primary School. https://doi.org/10.26914/c.cnkihy.2021.017508
[14] Chen, B. (2025). The logic, dilemmas, and practical paths of AI-empowered personalized oral English teaching. China Journal of Multimedia & Network Teaching (Mid-Monthly), (3), 87–90.
[15] Nguyen, H. A. (2024). Harnessing AI-based tools for enhancing English speaking proficiency: Impacts, challenges, and long-term engagement. International Journal of AI in Language Education, 1(2), 18-29.