Mechanisms of Digital Human-Driven Collaborative English Speaking Learning: An Interaction Hypothesis Perspective

Yang Yang*, Qiao Meng

International School, Guangzhou College of Commerce, Guangzhou 511363, China

Abstract: Grounded in the Interaction Hypothesis, this study investigates the mechanisms underlying digital human-driven collaborative English speaking learning. Addressing the persistent problem of interaction fossilization in computer-supported collaborative learning, the study argues that learners' low willingness to communicate and insufficient collaborative scaffolding often prevent peer interaction from developing into meaningful negotiation of meaning. Against this backdrop, digital human companions—characterized by multimodal interaction and enhanced social presence—are examined as potential mediators of interactional conditions conducive to second language acquisition. Drawing on interactionist theory, the study proposes a three-dimensional closed-loop interaction mechanism consisting of Pre-adaptation, Process Synergy, and Feedback Iteration. The pre-adaptation dimension calibrates learner profiles, task design, and affective readiness to establish optimal conditions for interaction. The process synergy dimension focuses on sustaining negotiation of meaning through scaffolded interaction, clarification requests, and prompts that encourage pushed output. The feedback iteration dimension embeds non-intrusive, data-driven corrective feedback and affective regulation into ongoing interaction, enabling continuous adjustment of task difficulty and interaction strategies. Theoretically, the proposed mechanism extends interactionist explanations of language learning to multimodal, human–AI collaborative environments. Practically, it offers a structured and implementable framework for designing intelligent speaking tasks that enhance interaction depth, feedback quality, and learner engagement. The study concludes that digital human companions can function as interactional facilitators rather than mere tools, supporting the transition from superficial participation to sustained, meaning-focused collaborative speaking.

Keywords: Digital Human Companions; Collaborative Speaking; Interaction Hypothesis; Negotiation of Meaning; Willingness to Communicate; Task-Based Language Learning

DOI: 10.57237/j.edu.2026.01.003

Funding: The Quality Engineering Construction Project of Guangzhou College of Commerce (Grant No. 2024JXGG49); The 2023 Guangdong Provincial Teaching Quality and Teaching Reform Project for Undergraduate Universities (Grant No. 2023JXGG07); The Guangzhou College of Commerce Project (Grant No. 2025XJYB010); The 2025 Higher Education Research Project under the "14th Five-Year Plan" of the Guangdong Higher Education Association (Grant No. 25GBY092); The 2025 China Association for Non-Government Education Planning Research Project (Grant No. CANQN250920).

*Corresponding author: Yang Yang, yyweixiao@qq.com

Received: 16 January 2026; Accepted: 30 January 2026; Published Online: 3 February 2026 http://www.educationrd.com

1 Introduction

In an era of increasingly complex global communication, English speaking instruction is undergoing a paradigm shift from "knowledge internalization" to "competence construction." Although Computer-Supported Collaborative Learning (CSCL) is widely recognized as an effective path to enhancing communicative competence, practical pedagogical settings often suffer from "interaction fossilization." This predicament, frequently triggered by learners' low Willingness to Communicate (WTC) [10] and a lack of collaborative scaffolding, prevents peer interactions from sparking deep negotiation of meaning.

With the evolution of Generative AI, "Digital Human Companions"—characterized by multimodal interaction and a high degree of social presence—offer a promising solution to break this deadlock [1, 3]. Rather than being mere static auxiliary tools, digital humans could serve as social agents and collaborative subjects. By simulating sophisticated social contexts, they provide a critical technological intervention point to alleviate speaking anxiety and reshape the interactive ecosystem.

However, while digital human technology has matured in terms of high-fidelity simulation, the internal mechanisms by which it drives language acquisition remain unclear. According to the Interaction Hypothesis, the core of language acquisition lies in obtaining comprehensible input through the negotiation of meaning, which in turn triggers learner output and corrective feedback [7, 11]. Current research, however, predominantly focuses on the technical implementation of digital humans or simple one-way dialogues. It often overlooks how digital humans, acting as "companions" within the collaborative process, can optimize interaction quality by intervening in the collaborative flow and providing corrective feedback. Consequently, how to precisely induce "negotiative behaviors" within digital human-driven collaborative settings and reveal the underlying support mechanisms for oral acquisition has become a cutting-edge challenge at the intersection of Educational Technology and Second Language Acquisition (SLA).

This study aims to systematically deconstruct the influence mechanisms of digital human-driven collaborative English speaking learning from the perspective of the Interaction Hypothesis. By constructing a specialized collaborative learning model, this research will explore the functional logic of digital human companions across dimensions such as enhancing interaction depth, optimizing feedback pathways, and increasing collaborative engagement. Theoretically, this study seeks to expand the connotations of human-AI collaborative learning and unveil language acquisition pathways under multimodal interaction. Practically, it aims to provide empirical evidence for designing intelligent speaking environments with high interactivity and emotional support, ultimately facilitating a substantive leap in English speaking instruction from "tool-driven empowerment" to "deep integration."

2 Literature Review

Digital human-driven collaborative English speaking learning refers to speaking-oriented activities in which learners interact with a human-like digital agent (e.g., an embodied conversational agent or virtual avatar) [1] while also collaborating with peers to complete communicative tasks. This topic can be grounded in long-established second language acquisition (SLA) research on interaction, because digital environments primarily matter insofar as they shape the quantity and quality of learner interaction. An Interaction Hypothesis perspective is therefore a conservative and well-supported way to explain potential learning mechanisms, focusing on how interaction can facilitate comprehensible input, negotiation of meaning, feedback, and opportunities for modified output [6, 8].

The Interaction Hypothesis proposes that language development is promoted when learners engage in interaction that helps them understand messages and resolve communication difficulties, often through conversational adjustments and repair sequences [8]. In task-based settings, communication problems can elicit clarification requests, confirmation checks, and comprehension checks, which are commonly labeled negotiation of meaning [4]. These interactional moves are relevant to speaking development because they link meaning-focused communication with moments of attention to form. A digital human-driven environment can be discussed within this established framework without assuming that the technology itself causes learning; rather, the central claim is that the digital human may help sustain interactional conditions known to be facilitative in SLA.

One interactional pathway concerns the improvement of input conditions during speaking tasks. Interaction research emphasizes that input becomes more useful when it is made

comprehensible and when it is modified in response to learners' needs during communication, rather than being simplified only in advance [6-8]. In a digital human setting, the agent can be designed to restate, paraphrase, slow down, or provide examples when learners signal misunderstanding. Such adjustments align with the interactionist notion of modified input and can occur in the context of meaningful dialogue. When learners collaborate in pairs or small groups, peers may jointly interpret the agent's prompts and negotiate how to respond, which can increase exposure to relevant language and create repeated encounters with key expressions within a task cycle.

A second pathway is negotiation of meaning and its relationship to noticing. The Noticing Hypothesis argues that learners must notice linguistic features in input for learning to occur [5, 12]. During negotiation episodes, learners' attention may be drawn to gaps in comprehension or production, making relevant forms more salient. Classic task-based studies have shown that communicative tasks can generate negotiation routines that support understanding and may promote attention to language form–meaning mappings. In digital human-driven interaction, the agent can prompt clarification or request more specific information when responses are ambiguous, potentially creating opportunities for noticing. Importantly, collaboration can make these noticing opportunities shared: learners discuss wording choices, attempt reformulations, and co-construct messages that the agent can understand, which keeps the activity meaning-focused while allowing brief shifts toward form.

Feedback and learner response constitute another well-established interactional mechanism. SLA research distinguishes various types of corrective feedback (e.g., recasts, prompts, explicit correction) and examines learner uptake, namely immediate responses that incorporate or react to feedback. Although findings vary across contexts and targets, a broad conclusion is that feedback can support development when it draws attention to form and when learners have opportunities to respond and self-repair [9]. In digital human contexts, feedback can be delivered in a consistent, trackable manner, such as through reformulations or clarification prompts embedded in dialogue. From a conservative standpoint, the relevant claim is not that digital feedback is inherently superior, but that it may increase the regularity of feedback episodes and the likelihood that learners encounter feedback tied to their own communicative attempts.

Speaking development also depends on opportunities for output. The Output Hypothesis argues that producing language can push learners to test hypotheses about form, notice gaps in their linguistic repertoire, and develop greater control over expression [14, 15]. In collaborative speaking, learners may engage in "languaging" or collaborative dialogue where they discuss and refine language while solving communicative problems, which can function as learning activity rather than mere practice [15]. A digital human can support this output-oriented process by maintaining task progression, prompting elaboration (e.g., asking for reasons or examples), and encouraging extended turns. These design elements fit naturally within task-based language teaching research, which emphasizes that task conditions shape interaction and performance, including fluency, accuracy, and complexity [2, 13].

Finally, affective conditions are relevant insofar as they influence participation in interaction. Speaking anxiety and fear of negative evaluation can reduce learners' willingness to communicate, limiting opportunities for interactional learning processes to occur. A digital human may be perceived as less judgmental, which may lower barriers to participation and increase speaking time. From an interactionist viewpoint, such affective benefits matter mainly because they can increase engagement in interaction—thereby increasing exposure to modified input, negotiation, feedback, and pushed output—rather than serving as a standalone mechanism.

3 Interaction Mechanism for Digital Human-Driven English Speaking Learning

This study constructs a three-dimensional closed-loop interaction mechanism consisting of "Pre-adaptation, Process Synergy, and Feedback Iteration." This mechanism deeply integrates the theoretical core of the Interaction Hypothesis with the multimodal interactive features of digital humans. By precisely addressing pragmatic negotiation pathways within a technological environment, it establishes a tripartite dynamic balance system encompassing "learner profile, task design, and interactive process," aimed at resolving the long-standing dilemmas of superficial interaction and low willingness to communicate in traditional oral collaboration.

As the logical starting point of the mechanism, the Pre-adaptation dimension focuses on the deep coupling of "data

profiling" and "situational modeling" to establish an optimal threshold for meaningful interaction. This dimension first leverages multimodal sensing technology to conduct a baseline assessment of the learner's pragmatic competence. This assessment extends beyond objective linguistic indicators such as vocabulary and grammar to include psychological thresholds regarding interaction frequency, turn-taking duration, and anxiety perception. Subsequently, based on the requirement for creating a "negotiation space" within the Interaction Hypothesis, the mechanism transforms static pedagogical goals into negotiatory tasks characterized by "communicative conflict." By pre-embedding "information gaps" and "cognitive contradictions" within the task pathway, it effectively induces proactive inquiry behaviors from the learner. Finally, through anthropomorphic introductory guidance, the digital human companion establishes its social identity as a "co-problem solver" rather than a "corrective examiner," thereby forging an egalitarian collaborative contract across both cognitive and psychological dimensions.

Table 1 Elemental Configuration and Functional Logic of the Pre-adaptation Dimension

Core Module	Interaction Hypothesis Mapping	Deepening Logic & Expected Objectives
Multidimensional Learner Diagnosis	Individual differences & input thresholds	To precisely establish the baseline of the learner's "Zone of Proximal Development", preventing cognitive overload or motivational loss caused by task mismatch.
Negotiatory Task Deconstruction	Information gaps inducing negotiation	To design negotiatory tasks featuring "decisional conflict" and "perspective complementarity," ensuring the interaction possesses the inherent momentum for meaning negotiation.
Consensus on Role Identity	Affective filter regulation mechanism	To establish a "Learning Community" through digital human strategies such as self-deprecation and empathy, thereby minimizing defensive psychological barriers during oral production.

As the central hub of the mechanism, the Process Synergy dimension focuses on transforming the interactive affordances of digital humans into scaffolding forces that drive learners to generate "pushed output." During the interaction initiation phase, the mechanism employs a "Situational Pulse" trigger mode, wherein the digital human initiates heuristic questions or challenges viewpoints in real-time based on collaborative progress to activate the learner's Willingness to Communicate (WTC). Upon entering the deep negotiation phase, the digital human companion utilizes Natural Language Processing (NLP) to diagnose pragmatic deviations. Rather than providing immediate standard answers, the mechanism employs strategies such as Clarification Requests and Confirmation Checks, compelling learners to reprocess linguistic forms through the "negotiation of meaning." This dynamic synergistic pathway not only achieves the leap from "comprehensible input" to "pushed output" but also ensures the continuity and cognitive depth of the collaborative interaction through the digital human's augmentation of viewpoints and logical extensions.

Table 2 Multi-Level Process Synergy Mechanism for Digital Human-Mediated Negotiation and Output Generation

Negotiation Level	Digital Human Intervention Strategy	Expected Learner Behavior	Interaction Outcome Goal
Trigger Level	Situational Pulse / Heuristic Prompting	Noticing / Initiation of Utterance	To awaken the Willingness to Communicate (WTC) and break interactive silence.
Negotiation Level	Clarification Request / Confirmation Check	Meaning Reconstruction / Linguistic Repair	To induce "pushed output" and enhance the accuracy of linguistic expression.
Deepening Level	Logical Extension / Argument Augmentation	Deep Argumentation / Co-construction	To expand interactive capacity and strengthen the depth of collaborative cognition.

The Feedback Iteration dimension serves as the critical support for the closed-loop mechanism, ensuring the long-term vitality of the interaction through the construction of a "whole-process data monitoring" and "non-intrusive corrective" system. This dimension breaks away from the singular summative assessment typical of traditional instruction by embedding feedback within every turn-taking exchange. Precision pragmatic feedback focuses on guiding learner self-correction through Recasts (implicit restructuring) while safeguarding the learner's sense of achievement. Simultaneously, the mechanism leverages affective computing to dynamically regulate the interaction rhythm; upon detecting frustration or peak anxiety, the digital human immediately switches to motivational strategies or provides more granular scaffolding. Ultimately, all interaction data is fed back into the learner profile in real-time, driving a

spiral optimization of task difficulty and interaction strategies for subsequent rounds, thereby achieving continuous iteration of the human-AI collaborative ecosystem.

Table 3 Multidimensional Matrix and Dynamic Iteration Logic of the Feedback Dimension

Feedback Dimension	Core Monitoring Indicators	Feedback Intervention Behavior	Mechanism Iteration & Evolution Logic
Pragmatic Cognitive Feedback	Lexical accuracy, turn duration, logical density	Recasts, prompted elicitation	Transforms feedback into "comprehensible input" for subsequent tasks, driving progressive improvement in linguistic accuracy.
Interactive Metacognition	Frequency of negotiation initiation, strategy success rate	Interaction reports, social strategy empowerment	Strengthens the learner's awareness of their own interaction process, facilitating a shift from "passive negotiation" to "proactive communication."
Affective Motivational Feedback	Emotional fluctuation curves, social presence scores	Empathic encouragement, interaction frequency adjustment	Establishes an emotional buffer zone and dynamically adjusts task load to ensure learners remain within a low-anxiety, active acquisition zone.

The entire mechanism adheres to a spiral logic of "Adaptation—Synergy—Iteration." Through the precise empowerment of the pre-adaptation dimension, the deep interaction of the process dimension, and the dynamic optimization of the feedback dimension, it achieves a substantive integration of the core tenets of the Interaction Hypothesis with the multimodal characteristics of digital humans. Theoretically, this mechanism expands the pragmatic explanatory space for human-AI collaborative learning; practically, it provides a quantifiable and intervenable paradigm for oral English instruction. By establishing a tripartite balance system encompassing "learner profile, task design, and interactive process," the digital human companion effectively compensates for the instability of peer support in traditional collaboration. Ultimately, this provides a robust mechanistic guarantee for high-quality English speaking acquisition within a digitalized context.

4 Conclusion

This study, grounded in the Interaction Hypothesis, systematically investigates the influence mechanisms of digital human-driven collaborative English speaking learning. The primary findings are as follows. First, by establishing a lifelike social presence, digital human companions effectively break the "interaction fossilization" common in traditional oral collaboration. The research indicates that digital humans act not only as information transmitters but also as regulators of the collaborative atmosphere. Their role as "companions" significantly lowers learners' affective filters, providing the psychological safety necessary for the negotiation of meaning. Second, the "Pre-adaptation—Process Synergy—Feedback Iteration" mechanism developed in this study successfully creates a logical loop from comprehensible input to pushed output. Through precise clarification requests and pragmatic restructuring, the digital human compels learners to reprocess linguistic forms, a deep cognitive engagement that directly promotes the internalization of speaking skills. Finally, this collaborative model shows a clear advantage in enhancing learners' Willingness to Communicate (WTC). Its non-intrusive feedback alleviates speaking anxiety and enables personalized learning empowered by technology.

Based on these conclusions, this study offers several implications for English speaking instruction in the era of digital transformation. At the pedagogical design level, the focus should shift from "results-oriented output" to the "process of interactive negotiation." Teachers should utilize digital humans to design collaborative tasks with "information gaps" to encourage higher-order language use. At the human-AI synergy level, the roles of digital humans and human teachers should be clearly defined: digital humans can provide high-frequency, routine scaffolding, while teachers focus on deep guidance and emotional support in complex social contexts. Furthermore, educational administrators should prioritize the personalized adaptation of digital human companions. By dynamically updating learner profiles, differentiated support can be provided to learners at various levels, making oral instruction both inclusive and precise.

Despite the progress in theoretical and mechanistic design, this study has limitations. On one hand, due to the limited duration of the empirical study, long-term tracking of language acquisition is still insufficient. Future research could conduct large-sample, longitudinal studies to explore the sustained impact of digital humans on linguistic growth. On the other hand, the accuracy of affective computing in digital humans can be further improved. Intervention strategies for the subtle emotional shifts learners experience during complex collaborations need to be more refined.

References

[1] Chen, S., Yang, L., Zhang, Y., Zhang, M., Xie, Y., Zhu, Z., & Li, J. (2025). Digital human technology in E-learning: Custom content solutions. *Applied Sciences*, *15*(7), 3807. https://doi.org/10.3390/app15073807

[2] Ellis, R. (2017). Task-based language teaching. In *The routledge handbook of instructed second language acquisition* (pp. 108–125). Routledge.

[3] Guo, P., Zhang, Q., Tian, C., Xue, W., & Feng, X. (2025). Digital human techniques for education reform. *Proceedings of the 2024 7th International Conference on Educational Technology Management*, 173–178. https://doi.org/10.1145/3711403.3711428

[4] Iwai, K. (2019). Designing tasks to encourage negotiation of meaning. *New Directions in Teaching and Learning English Discussion*, *7*, 126–132.

[5] Leow, R. P. (2018). Noticing hypothesis. *The TESOL Encyclopedia of English Language Teaching*, *1*(7). https://doi.org/10.1002/9781118784235.eelt0086

[6] Loewen, S., & Sato, M. (2018). Interaction and instructed second language acquisition. *Language Teaching*, *51*(3), 285–329. https://doi.org/10.1017/S0261444818000125

[7] Long, M. (1996). The role of the linguistic environment in second language acquisition. *Handbook of Second Language Acquisition*. https://cir.nii.ac.jp/crid/1571980075392115840

[8] Long, M. H. (1980). *Input, interaction, and second language acquisition*. University of California, Los Angeles.

[9] Lyster, R., & Mori, H. (2006). Interactional feedback and instructional counterbalance. *Studies in Second Language Acquisition*, *28*(2), 269–300. https://doi.org/10.1017/S0272263106060128

[10] Peng, J.-E., & Woodrow, L. (2010). Willingness to communicate in english: A model in the chinese EFL classroom context. *Language Learning*, *60*(4), 834–876. https://doi.org/10.1111/j.1467-9922.2010.00576.x

[11] Pica, T. (1994). Research on negotiation: What does it reveal about second-language learning conditions, processes, and outcomes? *Language Learning*, *44*(3), 493–527. https://doi.org/10.1111/j.1467-1770.1994.tb01115.x

[12] Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158.

[13] Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

[14] Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. *Input in Second Language Acquisition*, *15*, 165–179.

[15] Swain, M. (1995). Three functions of output in second language learning. *Principles and Practice in Applied Linguistics: Studies in Honor of HG Widdowson*, 125–144.