Innovative Practice of Artificial Intelligence-Assisted English Listening Teaching in Senior High School

Lili Zhang
The Middle School Attached to Ningxia University, Yinchuan, Ningxia, China

Abstract: Listening occupies 30 of the 150 points in the English paper of the National College Entrance Examination, yet listening lessons in many schools still follow a fixed routine of playing a recording, checking answers, and explaining the script. This paper reports an attempt to rebuild listening instruction around artificial intelligence at a senior high school in Yinchuan, Ningxia. It reviews the process view of listening and recent empirical studies on AI-supported listening and speaking instruction with adolescent and university learners, and it analyses three persistent problems in routine teaching: uniform input that ignores wide proficiency gaps, delayed and labour-heavy feedback, and low authenticity that feeds listening anxiety. A three-stage teaching model is then proposed, in which intelligent speech evaluation and chatbot tools handle diagnosis before class, multimodal interaction during class, and error-tagged review after class, while the teacher keeps control of task design and interpretation. One semester of practice in two parallel Grade 11 classes is described. The experimental class improved its mean score on a school-based listening test from 26.1 to 34.7 out of 50, against a rise from 26.0 to 29.5 in the parallel class, and 42 of its 52 students reported reduced listening anxiety. Conditions, risks, and boundaries of the practice are discussed.

Keywords: Artificial Intelligence; English Listening Teaching; Senior High School; Speech Recognition; Teaching Practice

1. Introduction

Listening occupies 30 of the 150 points in the English paper of the National College Entrance Examination, yet it rarely receives teaching design proportional to that weight. In a typical listening period at an ordinary senior high school, the teacher plays a recording twice, checks answers with the whole class, and reads through the script to explain difficult words. Students who already understood the material wait; students who understood almost nothing also wait, because the recording moves at one speed for everyone. After class there is no systematic record of which listener failed at which skill, so the next lesson starts from the same uniform point.

The situation persists partly for practical reasons. A teacher with two classes of around fifty students cannot transcribe individual listening errors, cannot prepare recordings at three difficulty levels every week, and cannot hold extended conversations in English with each learner. These are exactly the tasks at which machines have improved. Automatic speech recognition now scores pronunciation and dictation at sentence level within seconds [5], and chatbots built on large language models can hold topic-specific conversations, adjust wording on request, and present answers in text, audio, and image at the same time [4]. The question for a classroom teacher is no longer whether such tools exist but how to arrange them inside a coherent lesson structure without handing over the teaching itself.

This paper works on that question in the context of the Middle School Attached to Ningxia University in Yinchuan, where the author teaches English. Section 2 summarises the theoretical basis and the empirical evidence. Section 3 analyses the problems of current listening teaching practice. Section 4 presents a three-stage AI-assisted teaching model. Section 5 reports one semester of practice in two parallel classes, and Section 6 draws conclusions.

2. Theoretical Basis and Literature

2.1 Listening as an Active Process

Listening was long treated as a passive skill that improves simply through exposure. Field rejected this view and argued that comprehension breaks down at identifiable processing points, so teaching should diagnose

where decoding or meaning construction fails rather than re-play the whole recording [1]. Rost organised the research consensus around three questions, what listeners do before, during, and after listening, and placed strategy use at the centre of teaching [9]. Vandergrift and Goh turned this into a metacognitive pedagogy, in which students predict, verify, monitor, and evaluate their own comprehension across repeated listening rounds [2]. The instructional payoff has been tested. Graham and Macaro taught listening strategies to French learners of English over six months and reported clear gains in test performance compared with conventional instruction [3]. One implication matters here: effective listening teaching requires information about each listener's process, which is scarce in a class of fifty.

2.2 Artificial Intelligence in Listening and Speaking Instruction

Empirical work on AI-supported listening has grown quickly, and its findings are informative rather than uniformly positive. Tai and Chen assigned 88 ninth-grade learners in Taiwan to two experimental conditions using Google Assistant, one with a screen-based multimodal display and one with audio only, plus a conventional control group. After ten weeks, the multimodal group showed the largest gains in speaking proficiency, and interviews showed that the assistant served as a patient interlocutor that lowered anxiety [4]. Ji ran a one-year mixed-methods experiment in five Chinese middle schools using intelligent speech recognition and reported marked improvement in listening and speaking ability, together with regional differences in effect and persistent limits in recognition accuracy and semantic understanding [5]. Xiao conducted a randomised controlled trial with 84 university students in which the AI-driven speech recognition group improved listening comprehension and flow experience while listening anxiety fell, and the advantage was still present three weeks later [6]. Abdellatif and colleagues compared AI-based listening examinations with traditional ones among 57 learners and found gains not only in listening skill but also in self-competence and resilience [7]. A cautionary result comes from Zheldibayeva, whose 93 undergraduates using weekly chatbot activities outperformed the comparison group at the immediate post-test but not at the delayed follow-up, which suggests that short-term gains do not maintain themselves without continued structured practice [8]. The evidence therefore supports integration, not substitution.

3. Problems in Current Senior High School Listening Teaching

3.1 Uniform Input Against Wide Proficiency Gaps

A Grade 11 class in an ordinary school commonly contains students whose listening levels span three or more years of schooling. Some passed Cambridge-style tests in junior high; others can barely catch content words in a slow dialogue. The recording played in class, usually one examination-style passage, addresses none of them well. Strong students lose attention within minutes, weak students copy whatever neighbours write, and the middle group receives input only slightly above its comfort zone. The textbook workbook offers no branching difficulty, and the teacher has no time to prepare alternatives. Uniform input also hides the diagnostic information a teacher needs, because a wrong answer may come from unknown vocabulary, missed liaison, or absent background knowledge, and the answer sheet cannot tell these apart.

3.2 Delayed and Labour-Heavy Feedback

Feedback on listening arrives late and in coarse units. In most schools the answer sheet is marked for right and wrong only, sometimes by students exchanging papers. Where an individual error came from, whether a misheard vowel, a weak form, or a number heard in reverse, remains unknown unless the teacher interviews the student, which does not scale to a hundred students. Collecting and grading dictation weekly for two classes would take more preparation time than the timetable allows, so it is quietly dropped. The result is that listening instruction produces scores but not diagnoses, and the same students make the same errors across a whole semester.

3.3 Low Authenticity and Listening Anxiety

Classroom listening material is examination-shaped: standard accents, clean recordings, one-shot playback, and detached topics. Real communication is messier. It involves hesitations, reduced forms, varied accents, and the need to respond, none of which the routine

trains. For anxious students the one-shot playback adds pressure, since a missed sentence is gone and asking for a third replay marks one as slow. Research on listening anxiety has linked it to avoidance behaviour, including giving up on the listening section altogether in self-study time [6, 7]. A classroom with no interactive speaking-listening loop therefore trains a passive test posture rather than communicative listening.

3.4 Teacher Capacity and Routine Sustainability

A fourth difficulty receives less discussion: the sustainability of any attempt to change the routine. Listening periods are scheduled into tight examination preparation calendars, and a new routine competes with revision plans that parents and grade offices can measure in mock-test scores. Teachers also differ in their comfort with the tools themselves. In an informal survey of the author's teaching and research group, six of eleven colleagues had never opened a speech-evaluation platform, four had used one for examination scoring only, and one used error tags in lesson planning. The same capacity gap appears in the research literature: Ji found that the effect of speech-recognition technology varied with teachers' integration ability across the five schools studied, and listed teacher technical literacy among the conditions for success [5]. Any workable model must therefore survive a teacher who starts with minimal tool experience, must show benefits within one semester to maintain administrative support, and must reuse materials across weeks so that preparation effort falls rather than rises. These constraints shaped the design decisions described in the next section, in particular the fixed three-round structure and the reusable prompt and task library.

4. Design of the AI-Assisted Listening Teaching Model

4.1 Design Principles

Three principles guided the model. First, the teacher leads and the machine assists: every AI output enters the lesson through a task the teacher designed, and the teacher interprets the data before acting on it, following the human-in-the-loop stance recommended in recent implementation research [5, 8]. Second, diagnosis precedes instruction: the model collects individual error information continuously, so that whole-class teaching and individual review address measured weaknesses rather than assumed ones [1, 3]. Third, interaction replaces passive replay wherever the lesson goal allows it, because studies of assistant-based interaction attribute much of the benefit to the interlocutor role of the machine, including its patience, immediacy, and low-anxiety character [4].

4.2 The Three-Stage Workflow

Before class, students complete a ten-minute diagnostic micro-test on a speech-evaluation platform running an iFlytek-type recognition engine. The platform returns sentence-level scores and tags each error as misheard word, weak form, number or name, or inference failure, and pushes each student a personalised package: two re-listening clips targeting the tagged types plus five shadowing sentences at the student's measured level. The teacher reads the class-level error distribution in five minutes and adjusts the lesson focus accordingly.

During class, the weekly listening period is reorganised into three rounds. Round one is examination training with a difference: after the standard playback and checking, students scan their answer sheets and the platform shows which error types cluster where. Round two is assistant interaction. Working in pairs, students hold short task-specific conversations with a domestic large language model chatbot, such as ERNIE Bot or DeepSeek, by voice, asking for directions, booking, or interview answers, then compare the chatbot reply in text and audio with what they caught by ear. The multimodal display, text and audio arriving together, is kept deliberately, since multimodal feedback showed the strongest learning effects among the conditions tested by Tai and Chen [4]. Round three is note-taking and retelling against the same material at rising speed.

After class, each student keeps a listening journal on the platform. The system compiles the week's tagged errors into a review list and re-tests the same types in the following week, which implements the verify-and-monitor cycle of metacognitive pedagogy in machine-readable form [2]. The teacher checks journals weekly, not to grade them, but to catch students whose review lists stop moving.

4.3 Safeguards

Three boundaries are written into the routine.

Screen time during class interaction is capped at fifteen minutes per period. The platform's speech data is used only inside the class account and deleted at semester end, which reflects the data-privacy concern raised in school-based AI implementation [5]. Overreliance is checked by requiring students to retell content without replay before any second listening, so the machine supports practice rather than replaces it.

A further safeguard concerns equipment and equity. Not every student owns a smartphone suitable for the after-class platform, so the school keeps a set of thirty tablets borrowable from the English office for evening self-study, and the after-class tasks are also printed as a paper alternative with a QR code to the same recordings at the school library terminals. Grouping in the chatbot round is deliberate: each pair mixes a stronger and a weaker listener, and the pair changes every month, because pilot observation showed that two weak listeners in a pair often gave up the voice conversation and switched to typing, which removed the listening component of the task entirely. Table 1 summarises the workflow.

Table 1. The Three-Stage AI-Assisted Listening Teaching Workflow

Stage	Main Activities	AI Function	Teacher Role
Before class	Ten-minute diagnostic micro-test, personalised package	Speech evaluation, error tagging, material push	Read class error profile, adjust lesson focus
During class	Examination training, paired chatbot interaction, retelling	Voice conversation with text and audio display	Design tasks, interpret patterns, manage pacing
After class	Listening journal, weekly error review and re-test	Error list compilation, repeat testing	Check journals, identify stalled students

5. One Semester of Teaching Practice

5.1 Setting and Procedure

The practice ran through the autumn semester of 2025 in the author's school. Two parallel Grade 11 classes taught by the same teacher participated: Class 7 with 52 students used the model described in Section 4 in its weekly listening period plus after-class platform review, and Class 8 with 51 students continued the conventional routine with the same textbook and the same total listening time. Both classes sat a school-based listening test before and after the semester. The test contained five sub-parts, word recognition, detail comprehension, main idea, inference, and note-taking, each scored out of ten. The test was assembled by the teaching and research group from items used in previous school examinations, so that difficulty was comparable across forms, and both forms were trialled on a Grade 10 sample before use. At the end of the semester Class 7 completed a short anonymous questionnaire on anxiety and tool use, with six items on a five-point agreement scale and two open questions; the questionnaire drew on wording used in prior listening-anxiety research so that responses could be interpreted against published findings [6, 7].

5.2 Results and Observations

Class 7 rose from a mean of 26.1 to 34.7 out of 50; Class 8 rose from 26.0 to 29.5. The sub-skill profile was more informative than the totals. Detail comprehension, which the error-tagged review targeted most often, improved from 5.2 to 7.4 in Class 7 against 5.1 to 6.1 in Class 8. Note-taking improved from 4.1 to 6.0 against 4.2 to 4.8. Word recognition, already the strongest sub-skill, improved by 1.5 points in Class 7 and 0.5 in Class 8. Inference remained the weakest area in both classes, rising to 6.3 and 5.2 respectively, which matches its dependence on vocabulary and world knowledge rather than on listening practice alone.

The questionnaire returned 52 responses. Forty-two students agreed or strongly agreed that they felt less nervous when listening; nine were neutral; one disagreed. Thirty-seven reported using the chatbot interaction round as their favourite part of the lesson, and interview comments echoed the findings of earlier studies, describing the machine as patient and available for repeated attempts [4, 6]. Six students wrote that the machine sometimes misrecognised their pronunciation or produced replies too quickly, and two asked for more teacher explanation of difficult tags, a reminder that recognition accuracy and semantic depth remain limited [5].

The error-tagged logs also changed how the author prepared lessons. In week three, the aggregated log showed that 61 percent of logged errors in Class 7 were tagged as weak form or liaison, which the original lesson plan had not emphasised; a ten-minute mini-lesson on reduced forms was inserted into the following

period, and the share of that tag fell to 44 percent by week eight. Two students illustrate the individual value of the loop. One student, tagged repeatedly for numbers and names, was given a targeted package of train timetable and phone message recordings and stopped losing points on number items after four weeks. Another, whose tags spread evenly across types, turned out in conference to copy answers from a neighbour in the pre-test, a behaviour the micro-test made visible because it was taken individually on the platform. Neither case would have surfaced in the conventional routine of collected answer sheets.

Several cautions apply. The practice involved two classes, one school, and one semester, with no random assignment and a school-constructed test, so the numbers describe one teaching arrangement rather than a controlled estimate. The delayed follow-up question, whether gains persist after the platform is removed, remains open, since a one-semester advantage faded in prior university work once structured practice stopped [8]. The workload was real: preparing the task library for the chatbot rounds took the author about two extra hours per week in the first month, though this fell as prompts were reused.

6. Conclusion

This paper set out to answer a practical question: how artificial intelligence can be arranged inside senior high school listening teaching without dissolving the teacher's role. The answer proposed is a three-stage workflow in which the machine handles diagnosis, material matching, interaction, and error review, while the teacher designs tasks, interprets error profiles, and guards the boundaries of screen time, data use, and overreliance. One semester of practice in two parallel classes produced stronger growth in overall listening scores and in the sub-skills that error-tagged review targeted, and most students reported reduced anxiety, which is consistent with the pattern of findings from controlled studies of speech-recognition and chatbot-based instruction [4, 6, 7]. Two conclusions follow. First, the value of the tools lies less in their novelty than in the diagnostic loop they make possible, since knowing which error type each student repeats is what turns replay into teaching [1, 3]. Second, integration needs maintenance: schools adopting such a model should budget teacher preparation time, state data rules in advance, and plan how gains will be sustained once the structured routine is reduced [8]. Future work at the school will extend the model to Grade 12 revision and test whether the inference sub-skill responds to vocabulary-linked listening tasks.

References

[1] Field J. Listening in the Language Classroom. Cambridge: Cambridge University Press, 2008.

[2] Vandergrift L, Goh C C M. Teaching and Learning Second Language Listening: Metacognition in Action. London: Routledge, 2012.

[3] Graham S, Macaro E. Strategy instruction or listening comprehension: An intervention study. Language Teaching Research, 2008, 12(1): 49-74.

[4] Tai T Y, Chen H H J. The impact of intelligent personal assistants on adolescent EFL learners' speaking proficiency. Computer Assisted Language Learning, 2024, 37(5-6): 1224-1251.

[5] Ji H. Application of intelligent speech recognition technology in English listening and speaking teaching. International Journal of Web-Based Learning and Teaching Technologies, 2025, 20(1): 1-16.

[6] Xiao Y. The impact of AI-driven speech recognition on EFL listening comprehension, flow experience, and anxiety: A randomized controlled trial. Humanities and Social Sciences Communications, 2025, 12: 4672.

[7] Abdellatif M S, Alshehri M A, Alshehri H A, et al. I am all ears: Listening exams with AI and its traces on foreign language learners' mindsets, self-competence, resilience, and listening improvement. Language Testing in Asia, 2024, 14: 54.

[8] Zheldibayeva R. GenAI as a learning buddy for non-English majors: Effects on listening and writing performance. Educational Process: International Journal, 2025, 14: e2025051.

[9] Rost M. Teaching and Researching Listening. 3rd ed. London: Routledge, 2016.

