International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 2448-2457

https://doi.org/10.70517/ijhsa464207

A Study on the Transformation of English Teachers' Roles in Human-Machine Collaboration Modes

Wei Zheng[1,*]
[1] Foreign Language Department, School of Marxism, Huangshan Vocational and Technical College, Huangshan, Anhui, 245000, China
Corresponding authors: (e-mail: 13955979023@163.com).

Abstract This study explores the issue of role transformation for English teachers in human-machine collaboration models, with a particular focus on the application of speech synthesis technology in English teaching and its impact on teacher roles. A mixed-methods research approach was employed, combining questionnaire surveys and in-depth interviews to analyze data from 40 university English teachers and 200 students. Significant gaps were identified between expectations and reality regarding the modern teaching role enabled by technology. The expected proportion of teachers acting as “classroom activity facilitators” (87.5%) was significantly higher than the actual proportion (45%), and the expected proportion of teachers acting as “self-directed learners” exceeded the actual proportion by 24 percentage points. Regarding students' understanding of the transformation of teachers' roles, over 70% of respondents chose affirmative answers. However, in a survey on the specific content and methods of the transformation of teachers' roles, 49% of students indicated that they were not very familiar with the specific content and methods of the transformation. In terms of teachers' role transformation in the cognitive dimension, 65% of teachers believe that teachers should assume the role of student guides after the role transformation, while 47.5% of teachers believe that teachers should act as students' partners. The proportion of teachers who believe that teachers should remain in the dominant or authoritative role is zero.

Index Terms human-machine collaboration model, English teaching, speech synthesis technology, teacher role transformation

I. Introduction

In today's era of rapid technological advancement, the field of education has also witnessed significant transformations. Human-machine collaboration, as an emerging teaching model, is increasingly attracting attention and demonstrating unique advantages and potential in educational practice [1], [2]. Human-machine collaborative teaching refers to an interactive and collaborative model formed among teachers, students, and teaching robots or intelligent systems during the teaching process [3]. Human-machine collaboration is widely applied in English education. In traditional English teaching models, teachers are solely responsible for imparting knowledge and guiding student learning. In contrast, in human-machine collaborative teaching, teachers and intelligent systems jointly participate in teaching and guide student learning [4], [5]. The most significant feature of human-machine collaborative teaching lies in combining the strengths of teachers and intelligent systems to promote teaching innovation, providing students with more personalized, efficient, sustainable, and diverse learning experiences. This model plays a crucial role in alleviating student anxiety, promoting interaction, and optimizing resource allocation [6]-[9].

However, under the human-machine collaborative model, AI teaching assistants lead to a transformation in the role of teachers. Literature [10] explores how the teacher's role in the human-machine collaborative model shifts from “educator connecting with students” to “virtual teaching assistant + human teacher connecting with students,” with the virtual teaching assistant and human teacher complementing and collaborating to transform traditional English education. In this model, tasks such as homework grading, error correction, and feedback in English teaching are all automated. Literature [11] and [12] respectively developed an automatic scoring system for English essays and an intelligent scoring system for English translation, enabling rapid scoring while ensuring the objectivity, fairness, and quality of feedback. This avoids the subjectivity of teacher scoring and simplifies the teacher's workflow. A survey in Reference [13] showed that over 50% of students support using human-computer interaction on mobile devices for English teaching, integrating resources and providing teaching services through virtual scenario-based human-computer interaction. Reference [14] implemented human-computer interactive oral practice in online English speaking instruction supported by the Moodle platform, significantly improving students' speaking skills. Literature [15] notes that the English speaking exams for Chinese university students have transitioned from manual to computer-based testing, with AI-powered speaking examiners ensuring assessment accuracy. Literature [16] describes an improved convolutional neural network model designed to recognize students' facial emotions, assisting teachers in promptly monitoring students' psychological states and supporting online and remote English instruction. Literature [17] utilizes intelligent algorithms to assess students' incidental vocabulary acquisition, providing reference suggestions for vocabulary and reading, and assisting in teaching.

Additionally, the research in Literature [18] clarifies that while AI-driven English grammar assessment feedback tools possess a certain level of accuracy and effectiveness in understanding grammar rules, some students still rely on traditional feedback mechanisms, highlighting the necessity of a hybrid teaching model combining AI assistance with human teachers in human-machine collaboration. As the results of the survey in Reference [19] indicate, while AI-driven English instruction offers advantages such as immediate feedback, improved learning efficiency, and convenient access to resources, it cannot replace the cultural competence, critical thinking, and contextual understanding that human teachers provide to students. As pointed out in Reference [20], the anxiety caused by the transformation of the role of English teachers in AI-assisted teaching can be alleviated through three directions: cognitive restructuring, teaching innovation, and ethical repositioning, transforming them into “wise managers.” Therefore, it is necessary to explore the transformation of the role of English teachers in human-machine collaboration models to promote the sustainable development of human-machine collaboration in English education.

This paper first provides a detailed explanation of the speech signal preprocessing workflow, with a focus on key technical steps such as pre-emphasis, frame windowing, and endpoint detection. It introduces intelligent speech synthesis technology, compares the advantages and disadvantages of rule-based and data-driven methods, and designs a speech synthesis engine. Research on the transformation of English teachers' roles in human-machine collaboration modes is conducted, systematically demonstrating the complete process of using speech synthesis technology to assist English teaching. Qualitative research methods are employed, with a teacher-student paired design, to deeply explore teachers' adaptation processes in technology-enabled environments.

II. Voice synthesis technology design

II. A. Speech Signal Preprocessing

The first step in speech recognition is speech signal preprocessing. Speech signal preprocessing is the prerequisite and foundation of speech recognition, and it is also a critical step in extracting features from speech signals. Only by extracting feature parameters that represent the essence of the speech signal during the preprocessing stage can the best similarity results be achieved when comparing the speech signal with a standard speech signal. The speech signal preprocessing module typically includes five steps: speech signal digitization, endpoint detection, framing, windowing, and pre-emphasis.

(1) Speech Signal Digitization

Speech signals are time-varying waveforms and are analog signals. However, since computers only recognize digital signals, speech signals must be digitized for computers to process them. The process of digitizing speech signals involves sampling and quantization. After sampling and quantization, speech signals are converted into discrete digital signals.

(2) Pre-emphasis

The purpose of pre-emphasis is to enhance the high-frequency components of the speech signal and remove the low-frequency components, thereby flattening the signal spectrum. For speech signal spectra, higher frequencies typically have smaller amplitudes, and when the frequency of the speech signal doubles, the amplitude of the power spectrum decreases by 6 dB. To flatten the signal spectrum for easier analysis of spectral features, the speech signal must be pre-emphasized. High-frequency and low-frequency speech signals present different challenges in acquisition, and pre-emphasis is designed to address this contradiction. A common approach is to apply a digital filter with a 6 dB/octave high-frequency boost characteristic to the speech signal. This is a first-order digital filter:

$$H(z) = 1 - \mu * z^{-1} \tag{1}$$

If expressed in time domain form, the pre-emphasized signal $S_2(n)$ is:

$$S_2(n) = S(n) - *S(n-1) \tag{2}$$

In the formula, μ is set to 0.9375.

(3) Frame segmentation and windowing

Short-time analysis is used to frame the speech signal, meaning that the speech signal is divided into identical segments along the time axis. To ensure smooth transitions between adjacent frames and maintain continuity between adjacent frames, there must be an overlap between frames. This overlap is typically referred to as frame shift, and the amount of data contained in a single frame is called frame length.

To reduce the impact caused by the distance between two frames, a window function $w(n)$ is multiplied by the speech signal $s(n)$ during frame signal processing. Window functions generally use rectangular windows or Hamming windows, whose expressions are as follows:

Rectangular window:

$$W_R = \begin{cases} 1 & (0 \le n < N-1) \\ 0 & (Other) \end{cases} \tag{3}$$

Hanming Window:

$$W_{HAM} = \begin{cases} 0.5 - 0.46\cos(2\pi n/(N-1)) & (0 \le n < N-1) \\ 0 & (Other) \end{cases} \tag{4}$$

Han Ning Window:

$$W_{HN} = \begin{cases} 0.5 - 0.5\cos(2\pi n/(N-1)) & (0 \le n < N-1) \\ 0 & (Other) \end{cases} \tag{5}$$

The rectangular window provides spectral smoothing but results in the loss of waveform details; the Hamming window, on the other hand, effectively mitigates leakage and exhibits smooth low-pass characteristics. Therefore, in time-domain processing methods for speech, the rectangular window is typically selected, while in frequency-domain processing methods for speech, the Hamming window or Hanning window is generally chosen. Considering the different application ranges of rectangular windows and Hamming windows, the system uses a rectangular window function for endpoint detection of signals and a Hamming window for frequency domain transformation.

(4) Endpoint detection

Endpoint detection of speech refers to using computer digital processing technology to identify the starting and ending points of words and phrases in a segment of speech-containing signals, thereby storing and processing only the effective speech signals. Endpoint detection serves two important functions: first, it distinguishes between valid speech signals and silent or noisy signals, thereby removing a significant amount of useless signal data from the speech signal. Second, speech signals processed through endpoint detection can enhance feature extraction rates, thereby improving program runtime efficiency.

There is no fixed method for endpoint detection; different systems may employ different detection methods. The following introduces two commonly used methods.

1) Short-term energy

Short-term energy reflects the pattern of speech energy changes over time. Here, let the short-term energy of frame X_n of speech signal n be denoted by E_n. Then, its calculation formula is (where N is the frame length).

$$E_n = \sum_{m=0}^{N-1} x_n^2(m), 0 \le m \le N-1 \tag{6}$$

We can distinguish between speech and noise by analyzing the energy of the signal. Speech signals closer to the microphone exhibit higher energy. Using short-term energy, it is easy to distinguish between speech signals and noise signals in high signal-to-noise ratio environments. However, in low signal-to-noise ratio environments, short-term energy cannot clearly distinguish between speech signals and noise signals.

2) Short-term zero-crossing rate

The short-term zero-crossing rate refers to the number of times a short-term speech signal intersects the x-axis within one frame. The number of times a continuous speech signal crosses the time axis is the zero-crossing rate. For a discrete signal, if the signs of two consecutive sample values are different, it indicates one crossing of the time axis, from which the zero-crossing rate can be calculated. We define the short-term zero-crossing rate of a speech signal as:

$$Z_n = \frac{1}{2}\sum_{m=0}^{N-1} \left| \mathrm{sgn}[x_n(m)] - \mathrm{sgn}[x_n(m-1)] \right| \tag{7}$$

In the formula, $\mathrm{sgn}[x]$ is the sign function, i.e.

$$\mathrm{sgn}[x] = \begin{cases} 1 & (x \ge 0) \\ -1 & (x \le 0) \end{cases} \tag{8}$$

Low-energy, low-frequency sounds have a lower zero-crossing rate, while high-energy, high-frequency sounds have a higher zero-crossing rate. Based on an analysis of the characteristics of the Android platform and the functional requirements of the system, the system uses a combination of short-term zero-crossing rate and short-term energy analysis methods for endpoint detection. Both of these methods belong to the time domain analysis category and produce reliable and accurate results.

II. B.Intelligent speech synthesis technology

Speech synthesis technology, also known as text-to-speech (TTS) technology, is one of the core technologies used in the development of intelligent teaching assistants for English language learning. Selecting a high-performance TTS engine is therefore of utmost importance.

Rule-based and data-driven methods are the two main approaches to speech synthesis. Rule-based synthesis simulates the physical process of human speech production by defining rules, while data-driven synthesis relies on building a speech database to achieve speech synthesis. The specific classification of speech synthesis methods is shown in Figure 1.

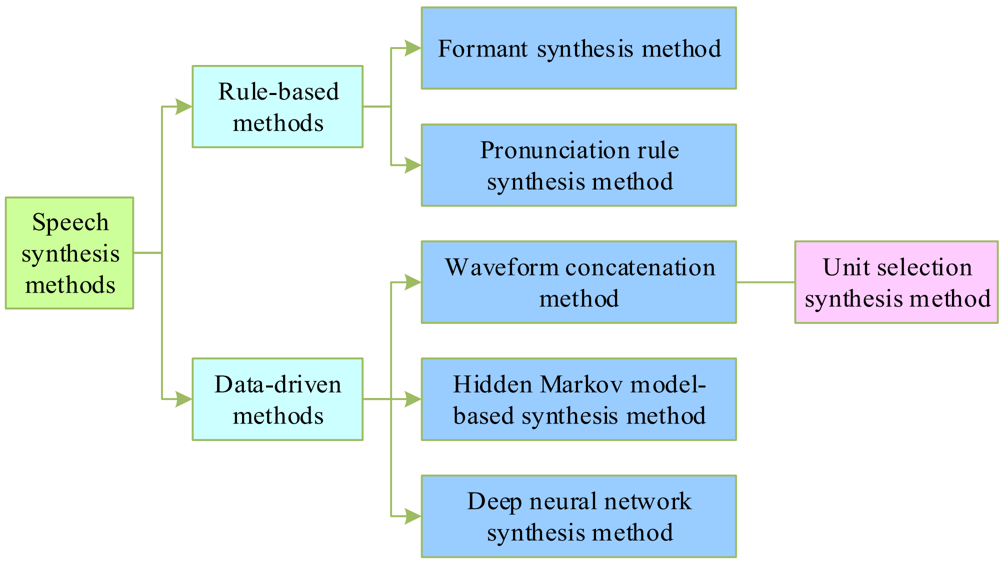


Figure 1: Classification of speech synthesis methods

Rule-based methods include formant synthesis and phonetic rule synthesis. Formant synthesis involves weighting and superimposing formants to generate speech, but speech synthesized using this method tends to sound unnatural. Phonetic rule synthesis directly simulates the human speech process by modeling a series of rules to generate speech. This method offers high flexibility and adaptability but results in speech that sounds overly mechanical and unnatural.

Data-driven methods primarily include hidden Markov model synthesis, waveform splicing synthesis, deep neural network synthesis, and unit synthesis. Waveform splicing involves slicing pre-recorded speech into smaller units, splicing these short units together, and then modifying them to synthesize speech content. This synthesis method maximizes the retention of human speech characteristics, offering high naturalness and clarity. However, the synthesized speech suffers from poor continuity, high error rates, and limited scalability. Unit selection synthesis can be viewed as an advanced form of waveform splicing. However, its speech database already contains a large number of prosodic examples, so the unit selection method eliminates the prosodic modification step in waveform splicing, thereby resolving the issue of discontinuity. Speech synthesized using this method has good sound quality, high naturalness, and stability, but it also has issues such as error-prone and poor scalability. The Hidden Markov Model (HMM) synthesis method partially addresses the challenges of large resource consumption and the need for highly detailed design in splicing methods. The HMM speech synthesis method saves system resources, offers good flexibility, and has low speech synthesis latency. However, the synthesized speech still lacks intonation variation. The deep neural network (DNN) synthesis method involves training a machine learning model using big data to establish a relationship model between text and its corresponding acoustic parameters, ultimately achieving speech synthesis. The DNN-based synthesis method has partially addressed the issues of traditional Markov methods, but there are still challenges in terms of accuracy.

Based on the design requirements of the intelligent teaching assistant and considering the current state of speech synthesis technology in terms of standardization, fluency, and naturalness, this paper designs a speech synthesis engine. This engine has the following advantages:

(1) The synthesized speech is standard, fluent, and natural, with high standards in terms of standardization and fluency, surpassing most English speech synthesis products currently available on the market. Additionally, this TTS engine has strong intonation variation capabilities, enabling it to switch between different intonations based on declarative or interrogative sentence structures, resulting in very high naturalness. This engine meets the standard

levels of English speech in terms of standardization, fluency, and naturalness, making it fully suitable for use as a teaching demonstration voice in educational settings.

(2) It features multiple voice libraries. In the English subject intelligent teaching assistant, a simple control interface allows for seamless switching between male and female voice libraries. The two voice library modes enable students of different genders to find suitable exemplary voices and also enable the intelligent teaching assistant to offer more functionalities (such as role-playing dialogue features). Therefore, teachers can use the intelligent teaching assistant to flexibly switch voice types, create dialogue-based materials, and conduct other teaching activities in the classroom.

(3) It has a complete voice library, with fast synthesis response and low latency. This TTS engine has a large voice library, which although it consumes more system resources, its extensive resource library enables the synthesized voice to be more natural and fluent. Additionally, its low synthesis latency and fast response speed enhance the system fluency of the intelligent teaching assistant and improve the user experience.

III. Research on the Transformation of English Teachers' Roles in Human-Machine Collaboration Modes

III. A. Study Design

With the rapid development of artificial intelligence technology, human-machine collaboration models are increasingly being applied in the field of education, posing new challenges to the traditional role of teachers. This study focuses on the application of speech synthesis technology in English language teaching and the resulting changes in the role of teachers.

III. A. 1) Method Selection

This study explores the application of speech synthesis technology in the field of education and delves into the issue of the evolving role of English teachers in human-machine collaboration models. As intelligent speech synthesis technology becomes increasingly prevalent in English instruction, teachers are transitioning from traditional knowledge disseminators to learning facilitators and technology coordinators. To comprehensively understand this transformation process, this research employs qualitative research methods, combining semi-structured interviews with focus group discussions to uncover the challenges and needs faced by teachers in human-machine collaboration settings. Semi-structured interviews help to understand teachers' subjective experiences, while focus group discussions facilitate the exchange of ideas among teachers, enabling feedback to be collected from multiple perspectives.

III. A. 2) Determination of research subjects

This study selected 40 English teachers from five different universities as research subjects, aiming to comprehensively understand the role transformation and professional development needs of English teachers in a digitalized environment. Participants were required to have a certain level of teaching experience and practical experience in using intelligent speech synthesis technology to assist in teaching, with usage durations ranging from six months to three years. The study subjects included teachers with varying teaching tenures, teaching backgrounds, and levels of technology application to ensure a multi-faceted reflection of teachers' diverse needs in human-machine collaboration models. By selecting a diverse group of teachers as subjects, the study can more comprehensively explore teachers' teaching practices and development needs in human-machine collaboration models.

To comprehensively assess the transformation of teachers' roles, the study also collected questionnaire data from 200 students, who were enrolled in the classes taught by the aforementioned teachers, ensuring data relevance. The student sample included 50 students from each of the first to fourth years, with a 1:1 male-to-female ratio and balanced English proficiency distribution (100% pass rate for CET-4 and 72% pass rate for CET-6). This teacher-student paired research design facilitates a multi-perspective analysis of the actual effects of human-machine collaboration models.

III. A. 3) Research Process

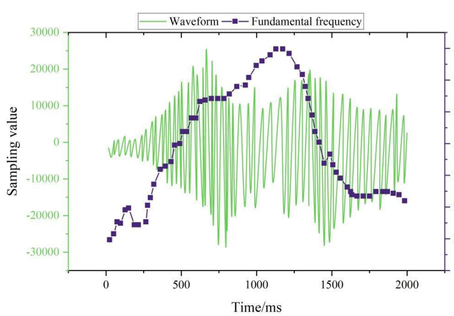
This study employed a mixed-methods approach, combining questionnaire surveys and in-depth interviews to collect data. The questionnaire was designed based on literature analysis and preliminary research findings, comprising 28 structured questions. It utilized a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree) and multiple-choice questions to measure teachers' perceptions, attitudes, and behavioral changes regarding role transitions. The questionnaire was distributed via the “QuestionStar” platform, achieving a 100% response rate. Prior to the formal interviews, three teachers were selected for exploratory interviews, which revealed that role transitions in human-machine collaboration modes involve teaching methods, psychological adaptation, content updates, and diverse student needs. Based on this, the research team revised the interview outline to focus on teachers' adaptation to human-machine collaboration modes, role transitions, professional development needs, and the challenges they face and their coping strategies. Data collection employed semi-structured interviews and focus group discussions to ensure participant privacy protection and data accuracy. All interviews were recorded and transcribed. Data analysis utilized thematic analysis to identify core themes related to teachers' role transitions, which were cross-validated through repeated dialogue analysis. This ultimately resulted in a framework for teachers' role transitions and professional development under the human-machine collaboration model. Theoretical saturation testing indicated that the data analysis had reached near saturation.

III. B. Teaching Process Assisted by Speech Synthesis Technology

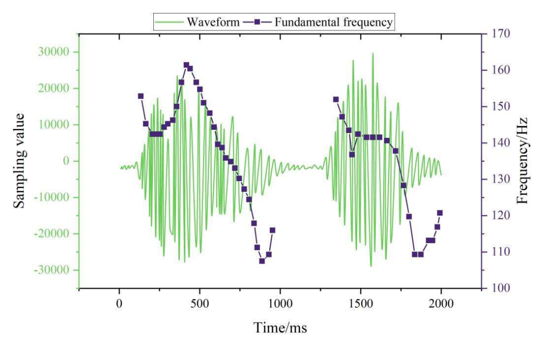
With the increasing application of intelligent speech synthesis technology in English education, teachers' teaching methods and role definitions are undergoing profound changes. Speech synthesis technology not only provides standardized models for pronunciation instruction but, more importantly, reconfigures traditional teacher-student interaction patterns through real-time feedback and personalized error correction functions. In this technological context, teachers are gradually transitioning from mere knowledge transmitters to designers of the learning process and guides for technological application. This section will focus on the specific applications of speech synthesis technology in two critical teaching phases: pronunciation correction and error detection. The speech text used for the experiment consists of 300 common English sentences, each read by 20 non-native speakers. All speakers possess university-level English proficiency and made efforts to avoid errors during reading. The reference standard speech uses the pronunciation of a native English speaker with an American accent. The acoustic model employs the speech synthesis engine proposed in this paper, which is based on dictation speech and is suitable for use as a scoring criterion.

III. B. 1) Corrective Feedback

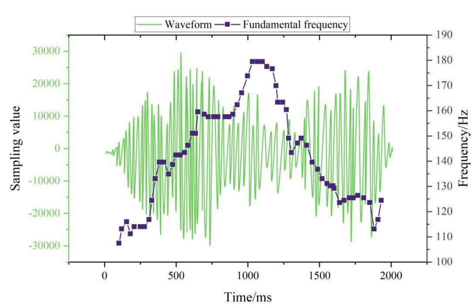
The speech waveforms of the fundamental frequency for the reference speech, learner speech, and prosodically corrected speech when a particular speaker reads “Where are” are shown in Figure 2 (a–c). It can be observed that the pause between “where” and “are” in the learner's speech is significantly longer than in the reference speech, reflecting the learner's inadequacy in linking techniques. The rhythmically corrected speech retains the main features of the original speech while optimizing the transition between words, making the speech flow more natural and coherent.



(a)Reference voice



(b)Learn pronunciation



(c)Corrected voice

Figure 2: Waveform fundamental frequency comparison of speech

III. B. 2) Error Detection

The statistical results for the least and most frequently mispronounced phonemes are shown in Table 1, with each row listing a mispronounced phoneme and the three phonemes it is most frequently mispronounced as. Error detection is effective for mispronounced or omitted phonemes. However, the error rate is high for inserted phonemes. In the error detection results of this validation corpus, there were a total of 637 mispronounced phoneme errors and 513 phoneme errors. From the statistical results, it can be seen that voiced consonants and complex vowels are prone to mispronunciation. This is partly due to the pronunciation habit of voicing in Mandarin Chinese, and partly because voiced consonants generally have shorter durations. Therefore, when providing corrective feedback, it is necessary to apply different tolerance thresholds for different phonemes.

Table 1: Statistical results for the least and most legible misphonemes

Misread the phoneme	Accuracy rate/%	Misreading as phonemes (misreading rate/%)
i	91.38	I(2.64), ʌ(1.75), eI(1.23)
k	89.62	t(3.11), p(1.98), g(1.29)
ʃ	83.45	z(4.62), s(2.31), dʒ(1.62)
ɔI	42.17	əu(8.45), ʌ(4.76), I(2.62)
au	36.12	ɔ:(13.15), a(5.22), ʌ(2.51)
tʃ	32.45	t(11.72), dʒ(4.66), d(1.17)
n	28.72	m(9.48), d(3.13), ŋ(1.25)

The most commonly elided phonemes are shown in Table 2. The elision rate for the plosive /d/ is the highest, reaching 15.23%, primarily occurring after consonants such as /t/ and /n/, which may be related to the pronunciation rule of incomplete plosives. The omission rates for the fricatives /ð/ and /v/ are 12.78% and 9.45%, respectively, and these sounds are often replaced by voiceless consonants with the same pronunciation. From a phonetic perspective, phoneme omission often occurs in complex consonant clusters or weak syllables, indicating that phonetic instruction should place particular emphasis on pronunciation training in these contexts.

Table 2: Most likely to skip phonemes

The phoneme was missed	Missed reading rate/%	Leading/trailing phoneme (missed reading rate/%)
d	15.23	t(6.42), n(4.35), l(2.16)
ð	12.78	θ(5.67), z(3.45), s(1.89)
v	9.45	f(4.23), w(2.15), b(1.07)
z	8.92	s(3.78), ʒ(2.14), dʒ(1.25)
θ	7.63	s(3.45), t(1.78), f(1.02)
ʒ	6.28	dʒ(2.87), ʃ(1.45), z(0.96)

III. C. Analysis of Teaching Models

III. C. 1) Questionnaire Results Analysis

The results of the student questionnaire survey clearly demonstrate the significant disparity between the expected and actual roles of English teachers in a human-machine collaboration model. For the two questions regarding “roles teachers have played” and “roles teachers expect to play,” the survey results are shown in Figure 3. The options “knowledge transmitter,” “course implementer,” “teaching activity organizer,” “teaching activity facilitator,” “daily learning monitor,” “teaching activity participant,” “classroom activity guide,” “self-directed learner,” and “knowledge system builder.” In terms of technology-enabled modern teaching roles, there is a notable gap between expectations and reality: the expected proportion of “classroom activity guides” (87.5%) is significantly higher than the actual proportion (45%), and the expected proportion of “self-directed learners” exceeds the actual proportion by 24 percentage points. The expected proportion of the “knowledge system builder” role differs from the actual proportion by 28 percentage points, reflecting teachers' need for transformation in course content reconstruction under human-machine collaboration models. Comparative analysis also reveals that the actual proportion of traditional managerial roles such as “daily learning monitor” is significantly higher than the expected proportion, while the expected proportion of interactive roles such as “participant in teaching activities” far exceeds the actual proportion. This contrast reflects the developmental trend of teachers' roles shifting from controllers to collaborators.

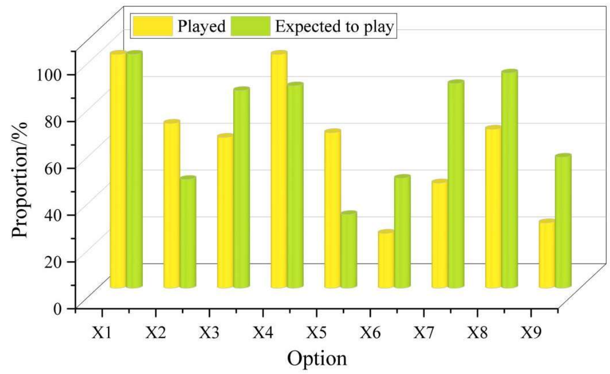


Figure 3: Results of the questionnaire survey

III. C. 2) Analysis of survey results

(1) Students' understanding of the transformation of the teacher's role

Students' understanding of the transformation of the teacher's role is shown in Table 3. In the survey on the extent of understanding of the phenomenon of the transformation of the teacher's role, the largest number of students chose the option "heard of it," accounting for 77%. The second largest number of students chose the option "not very clear," accounting for 17%, while the number of students who chose affirmative answers accounted for more than 70%. Students have a certain level of understanding of the phenomenon of teacher role transformation and are aware of its existence. However, their understanding of the specific content and methods of role transformation is not sufficiently in-depth, with some areas remaining unclear. In a survey on the specific content and methods of teacher role transformation, 49% of students indicated that they were not very familiar with the specific content and methods of teacher role transformation. This cognitive gap may be related to teachers not fully demonstrating the new role characteristics resulting from technology integration in actual teaching practice.

Table 3: Students' understanding of the changing role of teachers

Question	Option	Number of people	Percentage
Have you ever heard of the phenomenon of teachers' role transformation in schools?	Never heard of it	12	6%
	Not quite sure	34	17%
	Have heard of it	154	77%
Do you know the specific content and methods of the role transformation of teachers?	Have no idea at all	28	14%
	Out of line	98	49%
	Uncertain	24	12%
	Basic understanding	29	14.5%
	Have a complete understanding	21	10.5%

The statistical results of students' perceptions of the roles teachers assume after their role transformation are shown in Table 4. The highest number of respondents selected the options "organizer of teaching," "co-learner with students," and "family member and partner of students," with percentages of 64.5%, 59%, and 51.5%, respectively. This preference distribution reflects learners' expectations for collaborative teaching relationships in human-machine collaborative learning environments.

Table 4: Students' views on the roles played by teachers after their role transformation

Question	Option	Number of people	Percentage
What role do you think teachers should play after their role transformation?	The organizer of teaching	129	64.5%
	Co-learners of students	118	59%
	The student's family and friends	103	51.5%
	Student assessors and advisors	92	46%

	A coach and mentor for students	75	37.5%
	A growth counselor for students	61	30.5%
	Other	36	18%

As can be seen from the above, students are highly receptive to teachers undergoing role transformations and hold high expectations for such transformations, hoping to achieve their own development through these changes.

(2) Teachers' role transformations in the cognitive dimension

The statistical results regarding teachers' perceptions of their status after role transformations are shown in Table 5. 65% of teachers believe that teachers should assume the role of student guides after the role transformation, while 47.5% believe that teachers should act as students' partners. The proportion of teachers who believe they should remain in a dominant or authoritative role is zero. This indicates that the transformation of teachers' self-perception is highly correlated with the technical characteristics of human-machine collaboration environments. Intelligent voice technology has assumed some of the traditional functions of teachers, prompting them to redefine their own value.

Table 5: Teachers' views on their positions after the transformation of their roles

Question	Option	Number of people	Percentage
What position do you think teachers should be in after their role transformation?	Leader	0	0
	Authority	0	0
	Guide	26	65%
	Student partner	19	47.5%
	Other	13	32.5%

The statistical results of teachers' perceptions of normative requirements during role transitions are shown in Table 6. Teachers prioritize innovative teaching methods (92.5%), course design and development (80%), and guiding student development (67.5%) during role transitions. These norms and requirements place greater emphasis on innovation in teaching philosophy, methods, and content compared to traditional teacher role requirements, aligning with the new demands for teacher growth in contemporary educational environments. Establishing new normative requirements in teachers' cognition during the role transition process will help teachers clarify their rights and responsibilities after the role transition, thereby enabling them to make corresponding adjustments in their behavior.

Table 6: Teachers' perceptions of normative requirements during role transition

Question	Option	Number of people	Percentage
What normative requirements do you think should be followed to implement the role change of teachers?	Follow the law of education development	13	32.5%
	Abide by the teacher's code of conduct	11	27.5%
	Follow the code of ethics	9	22.5%
	Guide student development	27	67.5%
	Innovative teaching methods	37	92.5%
	Focus on curriculum design and development	32	80%
	Other	10	25%

IV. Conclusion

This study systematically explores the transformation of English teachers' roles in human-machine collaboration models, revealing the profound impact of speech synthesis technology on teaching practices and teacher-student relationships.

(1) In terms of technology-enabled modern teaching roles, there is a significant gap between expectations and reality: the expected proportion of “classroom activity facilitators” (87.5%) is significantly higher than the actual

proportion (45%), and the expected proportion of “independent learners” exceeds the actual proportion by 24 percentage points. The expected proportion of the “knowledge system builder” role differs from the actual proportion by 28 percentage points, reflecting teachers' need for transformation in course content reconstruction under human-machine collaboration models. Comparative analysis also reveals that the actual proportion of traditional managerial roles such as “daily learning monitor” is significantly higher than the expected proportion, while the expected proportion of interactive roles such as “teaching activity participant” far exceeds the actual proportion. This contrast reflects the developmental trend of teachers' roles shifting from controllers to collaborators.

(2) In terms of students' understanding of the transformation of teachers' roles, in a survey on the extent of understanding of the phenomenon of teacher role transformation, the highest proportion of students (77%) selected the option “heard about it.” The next highest proportion (17%) selected the option “not very clear,” while over 70% selected affirmative responses. In a survey on the specific content and methods of teacher role transformation, 49% of students indicated they were not very clear about the specific content and methods of teacher role transformation. The highest number of students selected the options “teacher as an organizer of teaching,” “teacher as a co-learner with students,” and “teacher as a family member and partner of students,” with percentages of 64.5%, 59%, and 51.5%, respectively.

In terms of the cognitive dimension of the transformation of the teacher's role, 65% of teachers believe that teachers should assume the role of student guides after the transformation, while 47.5% believe that teachers should act as students' partners. The proportion of teachers who believe that teachers should remain in a dominant or authoritative role is zero. During the role transformation process, teachers emphasize innovative teaching methods (92.5%), course design and development (80%), and guiding student development (67.5%).

References

[1] Meniado, J. C. (2024). Human–machine collaboration in language education in the age of artificial intelligence. RELC Journal, 55(2), 291-295.

[2] Nixdorf, S., Ansari, F., & Schlund, S. (2022). Reciprocal learning in human-machine collaboration: A multi-agent system framework in Industry 5.0. Digitization of the work environment for sustainable production, 207-225.

[3] Chi, J. (2022). Research on hybrid teaching model based on human-machine collaboration. BCP Educ. Psychol, 5, 19-24.

[4] Xuemei, W. (2025). DIGITAL DISCOURSE IN ENGLISH LANGUAGE TEACHING: EXPLORING MAN-MACHINE COLLABORATION MODELS THROUGH SYMBIOSIS THEORY. Journal of Pragmatics and Discourse Research, 5(2), 145-162.

[5] Qin, F., Sun, Q., Ye, Y., & Wang, L. (2025). A multimodal teaching model of college english based on human–computer interaction. International Journal of Human–Computer Interaction, 41(3), 1762-1770.

[6] Fu, J., Yan, H., & Zheng, J. (2023, December). An Empirical Study of Human-Machine Collaboration Teaching on English Listening Anxiety among Non-English Majors. In International Conference on Intelligent Information Hiding and Multimedia Signal Processing (pp. 1-11). Singapore: Springer Nature Singapore.

[7] Shang, H., & Sivaparthipan, C. B. (2022). Interactive teaching using human-machine interaction for higher education systems. Computers and Electrical Engineering, 100, 107811.

[8] Urlaub, P., & Dessein, E. (2022). From disrupted classrooms to human-machine collaboration? The pocket calculator, Google Translate, and the future of language education. L2 Journal: An electronic refereed journal for foreign and second language educators, 14(1).

[9] Zhao, L., & Yang, H. (2024). Mathematically Intelligent Human-Computer Collaborative Teaching: Opportunities, Challenges and Countermeasures. International Educational Research, 7(3), p11-p11.

[10] Wang, Z. (2025). Theories and Practices of Human-Machine Collab-oration Model in English Language Teaching: Ad-vantages, Challenges, and Future Development. The Educational Review, USA, 9(1), 92-97.

[11] Zhang, F. (2023). Design and application of an automatic scoring system for english composition based on artificial intelligence technology. International Journal of Advanced Computer Science and Applications, 14(8).

[12] Li, X., & Huang, C. (2025). Design of an intelligent grading system for college English translation based on big data technology. Systems and Soft Computing, 7, 200205.

[13] Chen, S., & Wang, J. (2024). Virtual reality human–computer interactive English education experience system based on mobile terminal. International Journal of Human–Computer Interaction, 40(13), 3560-3569.

[14] Ni, C. (2021, April). The human-computer interaction online oral English teaching mode based on Moodle platform. In 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC) (pp. 633-635). IEEE.

[15] Yu, Y., Han, L., Du, X., & Yu, J. (2022). An oral English evaluation model using artificial intelligence method. Mobile Information Systems, 2022(1), 3998886.

[16] Salloum, S. A., Alomari, K. M., Alfaisal, A. M., Aljanada, R. A., & Basiouni, A. (2025). Emotion recognition for enhanced learning: using AI to detect students’ emotions and adjust teaching methods. Smart Learning Environments, 12(1), 21.

[17] Zhang, W., & Wang, X. (2023). Intelligent Algorithm Evaluation of Incidental English Vocabulary Acquisition in Complex Reading Tasks. Journal of Electronics and Information Science, 8(3), 17-32.

[18] Nindya, M. A., & Taufiqulloh, T. (2024). EFL LEARNERS’PERCEIVED GRAMMAR ACCURACY: AI FEEDBACK TOOLS ASSESSMENT IN FOCUS. English Review: Journal of English Education, 12(3), 993-1002.

[19] Khan, I. A., Khan, K. A., & Mahmood, N. (2025). The Future of English Language Teaching in Pakistan: A Comparative Analysis of AI-Driven and Human-Led Instruction. Liberal Journal of Language & Literature Review, 3(1), 315-330.

[20] Jun, J. I. (2025). A. From Knowledge Authority to Wisdom Collaborator: The Transformation and Reconstruction of the Role of English Teachers in the Era of Gene AI. Journal of Literature and Art Studies, 15(3), 207-215.