DOI: 10.53469/jssh.2025.7(03).13

User Acceptance of Artificial Intelligence in University Oral English Instruction: An Analysis from the Perspective of Learners' Learning Styles

Yuan Fan

School of Foreign Studies, Lingnan Normal University, Zhanjiang, Guangdong, China

Abstract: ***Situated in an era when artificial intelligence has been applied across a wide range of fields, the domain of foreign language education has also gradually integrated new technologies. One representative technology applied in this field is Automatic Speech Recognition (ASR). The present study, based on the Unified Theory of Acceptance and Use of Technology (UTAUT) model, designs a user acceptance questionnaire for ASR and aims to explore the impact of learning styles on students' acceptance levels. The results demonstrate that visual, introverted, leveler and deductive learning styles are positively correlated with user acceptance of ASR; kinesthetic and sharpener learning styles are negatively linked to the user acceptance; while auditory and extraverted learning styles do not have significant impact on user acceptance of ASR. The results of this study provide important theoretical foundations and practical guidance for the application of ASR in foreign language education, particularly oral English instruction.***

Keywords: Automatic Speech Recognition (ASR), User acceptance, Learning styles, Oral English instruction.

1. Introduction

With the integration of technology and education globally, artificial intelligence (AI) and digital transformation are profoundly reshaping various fields. As a strategic technology in the new wave of technological revolution and industrial transformation, AI is rapidly becoming a key force driving humanity into the intelligent era (Rawas, 2024). In the domain of foreign language education, AI-assisted applications have been regarded as transformative tools, arousing learners' learning enthusiasm, and facilitating interactive learning process (Dandu et., 2024). Additionally, the incorporation of AI contributes to freeing language learners from physical and temporal classroom limits, allowing for easy access to multiple supportive learning resources (Hamuddin & Dahler, 2018). Diverse cutting-edge applications of AI in language teaching are employed to facilitate language learning and enhance learners' engagement, such as natural language processing, large language models, knowledge graphs, automatic speech recognition, intelligent robotics, and big data technology. Among these, Automatic Speech Recognition (ASR) technology, powered by deep learning neural networking, referring to a type of technology which can synchronously transcribe speech into texts (Shadiev et al., 2018). It has demonstrated unique advantages in oral English instruction due to its ability to comprehend and analyze language learners' speech, rendering it a valuable tool for second/foreign language learners, particularly in contexts where exposure to authentic, native language is limited (Dandu et., 2024). Additionally, ASR offers learners timely feedback and interactive oral exercises, creating a supportive, self-paced learning context (Chen et al., 2022). Given the ubiquitous use of mobile phones, the adoption of ASR-based applications in speaking courses is steadily increasing (Nguyen et al., 2018). The importance of the application of ASR in language teaching and learning, particularly on improving learners' speaking performance, is supported by a myriad of studies (e.g., Cavus & Ibrahim, 2017; de Vries et al., 2015; Nguyen et al., 2018; Tsai, 2023). However, while these studies tend to explore the effectiveness of ASR-based applications in improving learners' speaking performance or reducing speaking anxiety, they have primarily focused on the general trends and overlooked individual differences, such as the user acceptance of integrating ASR-based applications in language learning. To address this lacuna, this paper focuses on the influence of learners' preferred learning styles on their user experience of ASR-based applications, aiming to provide valuable references for the application and use of ASR-based tools in language teaching.

2. Basic Principles and Development Status of ASR Technology

ASR technology refers to the process of converting speech signals into corresponding written text or commands using computer-recognizable language (Michael, 2017). Unlike natural language processing (NLP), which analyzes and interprets language, ASR does not evaluate the semantics or coherence of language (Michael, 2017). Its basic principles involve matching input speech signals with pre-trained models through acoustic models, language models, and decoding algorithms to recognize and output corresponding text information. The acoustic model, the core component of an ASR system, converts speech signals into corresponding phoneme sequences. This model captures the acoustic features of speech signals through extensive training on large datasets, such as Hidden Markov Models (HMMs). Due to the advantages of Deep Neural Networks (DNNs) in nonlinear modeling, handling complex output layers, and adaptability to large datasets, DNNs are gradually replacing traditional HMMs, significantly improving the accuracy of speech recognition (Hinton et al., 2012). The language model predicts the probability of phoneme sequences forming words or sentences, helping the ASR system select the most linguistically appropriate output from multiple possible recognition results. In recent years, Neural Network Language Models (NNLMs) have gradually replaced earlier language models, such as statistical n-gram models, greatly enhancing the predictive capabilities of language models (Mikolov et al., 2013). The decoding algorithm combines the results of the acoustic and language models to generate the final recognized text. The most commonly used decoding algorithm is the Viterbi Algorithm, which uses dynamic programming to find the optimal path among possible speech signal paths, thereby outputting the best recognition result (Young et al., 2002).

With the advancement of computing power and big data technology, ASR technology has made significant progress. Early ASR systems were primarily used in specific fields, such as customer service automation systems and voice-controlled devices. In recent years, with the development of deep learning, ASR technology has broken through traditional limitations and is widely used in intelligent assistants (e.g., Apple's *Siri*, Huawei's *Xiaoyi*), online education, remote meetings, and smart home systems. In the field of foreign language education, ASR technology provides students with convenient tools for pronunciation practice and immediate oral feedback, effectively supporting language learners' independent learning. Additionally, ASR technology has made significant progress in handling regional accents, background noise, and multilingual recognition, further enhancing its application capabilities in education.

3. Application and Cases of ASR Technology in Assisting Oral English Instruction

The application of ASR technology in language learning has garnered widespread attention and research in recent years. Its core advantage lies in providing immediate speech feedback, helping learners better master pronunciation skills and improve oral expression abilities (Evers & Chen, 2021; McCrocklin, 2019). First, in pronunciation training, ASR technology can monitor and evaluate learners' pronunciation accuracy in real-time. Through various applications, ASR systems can "listen" to learners' pronunciation and provide formative assessments and feedback on pronunciation accuracy (Michael, 2017). This immediate feedback mechanism effectively helps learners to correct pronunciation issues at an early stage, preventing the formation of incorrect habits. Second, the application of ASR technology in classrooms is also extensive. For example, students can use tablets (often in pair activities) for oral or written tasks, such as simulated dialogues and role-playing. The ASR system can provide real-time corrections or assessments of learners' pronunciation or comprehensibility based on their speech input (Michael, 2017). This interactive learning method allows students to receive immediate corrections and evaluations during communication, thereby improving their oral skills and confidence. Furthermore, ASR technology provides learners with opportunities for independent learning (Yaniafari et al., 2022). Even without a conversation partner, learners can practice speaking through ASR technology. For example, students can read texts aloud or simulate dialogues independently, review and correct pronunciation errors by examining the text generated by the ASR system, and share their recordings with teachers for further guidance. This independent learning approach allows students to practice speaking anytime and anywhere, greatly enhancing the flexibility and autonomy of learning. Finally, ASR technology has shown significant advantages in automated oral assessment (Liu et al., 2019). It can automatically evaluate learners' oral performance and provide systematic assessment results. This automated assessment method reduces teachers' workload in grading and ensures the objectivity and consistency of evaluations. Automated assessment systems can evaluate multiple dimensions, such as pronunciation accuracy, fluency, and content completeness, helping learners comprehensively understand their oral abilities.

Currently, there are several mature intelligent oral assessment systems internationally, such as the Speech Rater developed by the Educational Testing Service (ETS), the Versant Speaking Test by Pearson, and other intelligent oral assessment platforms like the Duolingo English Test, Velawoods English, and Rosetta Stone. For example, Velawoods English, a self-study course jointly launched by Velawoods and Cambridge University Press, uses ASR technology to provide students with pronunciation and speaking feedback. The course features a game-like environment where learners can interact with virtual characters and simulate real-life dialogue scenarios (Michael, 2017). In China, there are also mature intelligent assessment systems, such as iFlytek's FiF Oral Training System and Chivox's Chinese-English Oral Assessment System. These systems use advanced ASR technology to provide accurate oral assessments and feedback, playing an important role in improving learners' oral abilities. For example, the FiF system uses color-coded feedback to indicate pronunciation accuracy: correctly pronounced words are marked in green, incorrect words in red, and moderately pronounced words in yellow. This visual feedback, combined with the system's scoring of pronunciation, fluency, and completeness, helps learners quickly identify and correct pronunciation issues, thereby improving their overall oral abilities.

Numerous studies have been conducted on the application of ASR technology in improving learners' pronunciation accuracy and fluency, enhancing learning autonomy, and reducing teachers' repetitive workloads. For example, Elimat et al. (2014) found that ASR technology significantly improved the average pronunciation scores of experimental group students, especially during independent practice sessions. ASR technology can also provide learners with a more relaxed and low-pressure learning environment. For instance, Chiu et al. (2007) found that using ASR technology for oral practice reduced students' anxiety about making pronunciation errors in class, thereby increasing their confidence in speaking practice. This finding aligns with Bashori et al.'s (2021) study, which revealed that students who received ASR-based interventions experienced less anxiety and greater learning enjoyment compared to the control group, who attended regular classes. McCrocklin (2016) showed that ASR technology provides a rich English environment with error recognition and feedback, helping students monitor and correct pronunciation errors, thereby promoting their autonomous learning and improving independent learning abilities. Additionally, Jiang et al. (2021) pointed out that ASR technology can reduce teachers' repetitive work in checking students' pronunciation and morphosyntactic errors. However, research on the adaptability of ASR technology to different learning styles and its acceptance among learners is relatively limited. As Griffiths and Soruç (2020) noted, acknowledging individual differences can enhance learners' enjoyment and learning outcomes, and a "one-size-fits-all" teaching approach should be avoided. Therefore, it is necessary to explore the adaptability of ASR technology to different learning styles and the impact of these styles on ASR acceptance. This will help foreign language learning platform developers design better ASR platforms and assist foreign language teachers and learners in utilizing ASR technology more effectively, thereby promoting the diversification and personalization of foreign language teaching methods.

4. Research Design

An ASR technology acceptance questionnaire, based on the Unified Theory of Acceptance and Use of Technology (UTAUT) model, was designed to evaluate learners' acceptance of ASR-based mobile foreign language learning platforms, with a focus on college students. The goal of this study is to provide suggestions for improving the design and application of ASR-supported mobile English language learning resources and to promote the diversification of English pronunciation teaching in universities. The UTAUT model, proposed by Venkatesh et al. (2003), integrates previous technology acceptance research and forms a comprehensive scale (Venkatesh et al., 2012). This study adopts five key dimensions from the UTAUT model: performance expectancy, effort expectancy, social influence, facilitating conditions, and hedonic motivation. The original questionnaire items were modified by replacing "mobile internet" with "speech recognition technology". Specifically, performance expectancy refers to the benefits that using ASR technology can bring to learners in specific activities; effort expectancy refers to the convenience of using ASR technology; social influence refers to the degree to which learners perceive that important others (e.g., peers and teachers) believe they should use ASR technology; facilitating conditions refer to learners' perceptions of the resources and support needed to use ASR technology; and hedonic motivation refers to the pleasure or enjoyment learners derive from using ASR technology, which has been proven to play a crucial role in user acceptance of technology (Brown & Venkatesh, 2005). The specific dimensions of user acceptance of ASR-based mobile learning are shown in Table 1. Additionally, individual difference variables, such as learning styles, are considered to moderate these relationships (Venkatesh et al., 2012).

Table 1: Key dimensions of user acceptance of ASR-based mobile learning

Dimension	Definition
Performance expectancy	Learners' belief that using ASR technology helps improve their learning efficiency
Effort expectancy	Learners' perception of the convenience of using ASR technology
Social influence	Learners' perception of important others' belief that they should use ASR technology
Facilitating conditions	Learners' perception of resources and support provided by ASR technology
Hedonic Motivation	Learners' enjoyment or pleasure derived from using ASR technology

Therefore, this study uses the Learning Style Survey (LSS) to measure students' learning styles and determine whether learning styles affect their acceptance of ASR technology. The LSS, developed by Cohen et al. (2001), assesses individuals' learning preferences. This study selects four dimensions related to ASR technology use: visual-auditory-kinesthetic perceptual styles, extraverted- introverted personality types, sharpener-leveler information processing styles, and deductive-inductive reasoning styles. These four key dimensions belong to three categories of learning styles: perceptual learning styles, personality learning styles, and cognitive learning styles. The specific learning style types, dimensions, and focuses are shown in Table 2. These dimensions were chosen because they directly influence how students receive and process information, thereby affecting their acceptance of ASR technology. First, the visual-auditory-kinesthetic perceptual styles help identify which senses students prefer to use during learning, which is crucial for the audio and visual feedback provided by ASR technology. Second, the extraverted-introverted personality types reveal students' learning preferences in social environments. Extraverted students may prefer interacting with others through ASR, while introverted students may prefer independent learning. Third, the sharpener-leveler information processing styles indicate whether students focus on specific details or overall understanding during learning, which is closely related to the personalized feedback provided by ASR technology. Finally, the deductive-inductive reasoning styles show students' tendencies in understanding language rules. Deductive learners may prefer learning specific rules through ASR, while inductive learners may prefer summarizing rules from examples. Analyzing these dimensions will provide valuable insights for ASR product developers and educators, revealing how to optimize ASR applications based on students' learning styles and enhance their effectiveness in oral English instruction.

Table 2: Learning style types, dimensions, and focuses

Category	Dimension	Focus
Perceptual Learning Styles	Visual-Auditory-Kinesthetic	How learners use their senses to learn
Personality Learning Styles	Extraverted-Introverted	Learners' preference for self or group learning
Cognitive Learning Styles	Sharpener-Leveler	How learners memorize new learning materials
	Deductive-Inductive	How learners understand language rules

The ASR technology perception questionnaire and the Learning Style Survey used in this study are both in Chinese and have been back-translated to ensure the accuracy and consistency of the questionnaire content. Through these measures, this study aims to more accurately assess Chinese college students' learning styles and their acceptance of ASR technology.

5. Results and Discussion

5.1 Research Participants

The questionnaires were distributed through the online platform "*Wenjuanxing*" and provided rewards to participants upon completion. A total of 347 questionnaires were collected from a university in southern China, of which 32 were excluded due to invalidity (e.g., all questions answered with the same response), leaving 315 valid questionnaires for data analysis. The results show that the male-to-female ratio in the valid sample was 4:6. Among the participants, the proportions of first- to fourth-year undergraduates were 30.4%, 28.8%, 29.5%, and 11.3%, respectively. Additionally, 93% of participants reported using ASR-based language learning applications more than once a month.

5.2 Reliability Analysis

Cronbach's alpha is an important indicator for assessing questionnaire reliability and is widely used in empirical data analysis. This study uses Corrected Item-Total Correlation (CITC) to evaluate the reliability of each questionnaire item. The Cronbach's alpha coefficients for each dimension range from 0.787 to 0.962, and all variables have Cronbach's alpha coefficients greater than 0.7. The CITC values and Cronbach's alpha coefficients after item deletion also meet research requirements. This indicates high stability of the variables in the questionnaire, meaning the items have strong homogeneity.

5.3 Validity and Factor Analysis

The results of the Kaiser-Meyer-Olkin (KMO) test and Bartlett's test of sphericity are shown in Table 3. Factor analysis is used for information condensation, but before conducting it, it is necessary to assess whether the data are suitable for factor analysis. The results show that the KMO value is 0.929, which is greater than 0.7, meeting the prerequisite for factor analysis. This means the data can be used for factor analysis. Additionally, the data passed Bartlett's test of sphericity ($p < 0.05$), indicating that the data are suitable for factor analysis. This study uses the varimax rotation method to determine the correspondence between factors and items. The results show that all items have communality values higher than 0.4, and the absolute values of the corresponding factor loadings are greater than 0.5, indicating strong associations between items and factors. Therefore, the scale has good structural validity.

Table 3: KMO and Bartlett's Test results

KMO		0.929
Bartlett's Test of Sphericity	χ^2	14820.306
	df	2485
	sig	0.000

5.4 Confirmatory Factor Analysis

This study uses AMOS 26.0 software to conduct confirmatory factor analysis (CFA) on variables such as visual, auditory, kinesthetic, extraverted, introverted, sharpener, leveler, deductive, inductive, and user acceptance. Additionally, composite reliability (CR) and average variance extracted (AVE) are used to assess the convergent validity of each variable dimension. First, the goodness-of-fit of the CFA model was tested. The data collected from the questionnaire were imported into AMOS 26.0, and the model fit parameters obtained using the maximum likelihood method are shown in Table 4.

Table 4: Model fit for confirmatory factor analysis

Model Fit	CMIN	DF	CMIN/DF	NFI	RFI	IFI	TLI	CFI	GFI	RMSEA
Fit Results	2843.649	2369.000	1.200	0.824	0.815	0.965	0.963	0.965	0.810	0.025
Judgment Value			<3	>0.8	>0.8	>0.9	>0.9	>0.9	>0.8	<0.08

The results show that the CMIN/DF (chi-square minimum/degrees of freedom) value is 1.200, which is less than 3, indicating a good model fit. Additionally, the NFI (normed fit index), RFI (relative fit index), IFI (incremental fit index), TLI (Tucker-Lewis index), CFI (comparative fit index), and GFI (goodness-of-fit index) all reached excellent levels, and the RMSEA (root mean square error of approximation) is 0.025, which is less than 0.08, indicating a good model fit. Overall, the model fit for each variable is good, and the confirmatory factor analysis is valid.

5.5 Structural Equation Model Analysis

Table 5: Structural model path analysis results

Path	Estimate	S.E.	C.R.	p	β
Acceptance ← Visual	0.167	0.069	2.4	0.016	0.154
Acceptance ← Auditory	0.035	0.069	0.502	0.616	0.027
Acceptance ← Kinesthetic	-0.285	0.084	-3.376	<0.001	-0.212
Acceptance ← Extraverted	-0.072	0.053	-1.353	0.176	-0.073
Acceptance ← Introverted	0.152	0.062	2.446	0.014	0.15
Acceptance ← Sharpener	-0.155	0.066	-2.366	0.018	-0.144
Acceptance ← Leveler	0.163	0.058	2.798	0.005	0.166
Acceptance ← Deductive	0.193	0.076	2.55	0.011	0.174
Acceptance ← Inductive	0.059	0.107	0.546	0.585	0.045

Using AMOS 26.0, this study added latent variables such as visual, auditory, kinesthetic, extraverted, introverted, sharpener, leveler, deductive, inductive, and user acceptance. First, based on the research model's assumptions about the relationships between latent variables, a structural equation model framework was constructed. Next, based on the items in the scale, measurable variables and corresponding residual variables were set for the latent variables. The final structural model path analysis results are shown in Figure 5. The results indicate that visual learning style has a positive impact on user acceptance ($\beta = 0.154$, $p = 0.016$), possibly because visual learners prefer learning through images and visuals, making them more receptive to the textual feedback provided by ASR technology. Auditory learning style does not significantly affect acceptance ($\beta = 0.027$, $p = 0.616$), suggesting that ASR applications may not meet the learning needs of auditory learners, and more interactive and multimodal feedback (e.g., video or audio) may be needed to enhance their acceptance. Kinesthetic learning style has a negative impact on acceptance ($\beta = -0.212$, $p < 0.001$), possibly because kinesthetic learners prefer learning through hands-on practice and may find purely ASR-based oral exercises less engaging or challenging. Extraverted personality does not significantly affect acceptance ($\beta = -0.073$, $p = 0.176$), possibly because extraverted learners require more socially oriented tasks to facilitate learning. Introverted personality has a positive impact on acceptance ($\beta = 0.15$, $p = 0.014$), likely because introverted learners prefer independent learning, and ASR technology provides personalized feedback, allowing them to learn in a relatively quiet environment while reducing anxiety associated with real-life conversations. Sharpener learning style has a negative impact on acceptance ($\beta = -0.144$, $p = 0.018$), possibly because sharpeners focus on details, and ASR-based platforms may not provide sufficiently detailed feedback, leading to lower acceptance. Leveler learning style has a positive impact on acceptance ($\beta = 0.166$, $p = 0.005$), likely because levelers prefer understanding overall concepts, and ASR-based platforms help them improve oral expression through holistic language exercises. Deductive learning style has a positive impact on acceptance ($\beta = 0.174$, $p = 0.011$), indicating that ASR applications that provide learning rules first align with deductive learners' preferences. Inductive learning style does not significantly affect acceptance (β = 0.045, p = 0.585), possibly because ASR-based learning platforms do not provide rich contexts for inductive learners to autonomously summarize rules, resulting in moderate acceptance.

6. Conclusion and Recommendations

This study analyzed the influence of different learning styles on the acceptance of ASR technology using structural equation modeling. The results show that visual, introverted, leveler, and deductive learning styles are positively correlated with ASR acceptance, while kinesthetic and sharpener learning styles are negatively correlated. Auditory and extraverted learning styles do not significantly affect ASR acceptance. These findings are closely related to the learning content and methods provided by existing ASR-based oral learning platforms or applications. This study reveals that individual differences among learners significantly affect their acceptance of ASR technology. Therefore, understanding these differences provides important references for the design of ASR technology and the effective use of this resource by language instructors.

6.1 Recommendations for ASR-based Oral Learning Platform Design

To meet the needs of different learning styles, ASR platforms should provide diverse and multimodal oral practice and feedback mechanisms, allowing users to customize platform settings based on their learning styles and preferences. For example, users should be able to choose learning content (visual, audio, or hands-on activities; independent or interactive; rules-first or examples-first) and feedback types (detailed or holistic feedback) to enhance motivation and learning outcomes.

For visual learners: ASR platforms should provide rich visual materials, such as figures, video demonstrations, and dynamic feedback, to help users better understand pronunciation and grammar structures. Visual aids can be integrated into oral exercises to make the learning process more intuitive. For auditory learners: ASR platforms should focus on providing high-quality audio materials, including clear speech demonstrations and dialogue practice, encouraging learners to imitate and repeat. Additionally, platforms should offer audio comparison features, allowing learners to compare their recordings with standard pronunciations and receive specific improvement suggestions. For kinesthetic learners: ASR platforms should design highly interactive activities, such as role-playing or scenario simulations, allowing learners to practice language through hands-on activities. Incorporating virtual reality (VR) and augmented reality (AR) technologies can create immersive learning environments, enhancing engagement and motivation. For example, learners can enter virtual scenarios through VR to interact with virtual characters or use AR to combine real-world environments with language learning tasks, rendering the experience more dynamic and practical. Additionally, platforms can integrate an instant online dialogue feature, using visual interfaces such as online user lists, status indicators, and interest tags to help learners quickly find other online users for real-time interaction, topic discussions, or collaborative tasks. This feature not only enhances the social aspect of learning but also provides kinesthetic learners with more opportunities to practice language, further improving learning outcomes.

The real-time interaction can be also applied to extraverted learners, for example, ASR platforms should provide group activities and real-time conversation opportunities, allowing users to randomly match with conversation partners for interactive practice. Cooperative tasks can also be designed to encourage learners to improve their oral expression and social skills through interaction. For introverted learners, ASR platforms should offer more human-computer interaction and immediate feedback, reducing anxiety associated with interacting with others while enabling independent learning and gradual mastery of language skills.

For sharpener learners who show more inclination for specific details, ASR platforms should provide detailed feedback and link to specific language exercises, helping learners focus on details during the learning process. Targeted exercises can improve their language accuracy and enhance their engagement in the ASR-based exercises. As for leveler learners, ASR platforms should focus on understanding overall concepts, providing integrated learning tasks that allow learners to improve oral expression within a broader context. Scenario-based exercises can help them understand the macro framework of language use.

For deductive learners, ASR platforms should first provide rules and theories, followed by examples for application. Exercises should be designed to help learners flexibly apply the rules they have learned. In contrast, for inductive learners, ASR platforms should first provide rich examples and scenarios, allowing learners to summarize language rules through practice. Guiding learners from concrete to abstract learning can effectively enhance their interest and mastery of language rules.

6.2 Recommendations for Language Teachers Using ASR Technology

With the increasing prevalence of advanced technology such as ASR, language teachers can effectively integrate this technology into oral instruction to enhance students' learning outcomes and engagement. For example, teachers can design classroom activities that incorporate various ASR-based interactive tasks, such as story continuation, simulated dialogues, oral essays, and peer evaluations. The immediate feedback provided by ASR can visually transform students' oral expressions into text, promoting collaboration and communication among students while helping them identify and correct errors in their speech, thereby improving their oral expression skills.

Additionally, teachers can combine ASR technology with text, audio, and video resources to provide multimodal learning experiences that cater to different learning style preferences. This approach not only meets the needs of diverse learners but also enhances students' understanding and retention of learning content. Where possible, teachers can use ASR platforms integrated with AR or VR technologies to create realistic scenarios for language practice, further enhancing the immersive learning experience and students' acceptance of ASR technology.

Outside the classroom, teachers should encourage students to use ASR platforms for independent oral practice, tailoring learning paths and practice content to different types of learners. By offering optional learning modules, students can practice based on their interests and needs, thereby increasing motivation. Teachers should also guide students to make full use of the immediate feedback provided by ASR technology, encouraging them to step out of their comfort zones and adapt to different teaching styles to improve their language abilities.

6.3 Limitations and Future Research

This study has some limitations. First, the sample size is relatively small and limited to one university, which may restrict the generalizability of the results. Additionally, this study only focused on four learning styles, leaving out other styles that may affect ASR acceptance. Future research should expand the sample size, include more learning styles, and explore the potential of ASR technology in independent learning outside the classroom and its integration with emerging technologies like AR and VR to more comprehensively enhance language learning outcomes.

Acknowledgments

This research was supported by the following fund: Guangdong Higher Education Teaching Reform Project 2023 No.4 (YueJiaoGaoHan [2023] No.4): "Research and Reform Practice on Improving College English Students' Pronunciation Accuracy Based on ASR Technology"

References

[1] Bashori, M., van Hout, R., Strik, H., & Cucchiarini, C. (2021). Effects of ASR-based websites on EFL learners' vocabulary, speaking anxiety, and language enjoyment. *System*, 99, 102496. https://doi.org/10.1016/j.system.2021.102496

[2] Brown, S. A., & Venkatesh, V. (2005). Model of adoption of technology in households: A baseline model test and extension incorporating household life cycle. *MIS Quarterly*, 29(3), 399–426. https://doi.org/10.2307/25148690

[3] Cavus, N., & Ibrahim, D. (2017). Learning English using children's stories in mobile devices. *British Journal of Educational Technology*, 48(2), 625–641. https://doi.org/10.1111/bjet.12427

[4] Chen, C.H., Koong, C.S., & Liao, C. (2022). Influences of integrating dynamic assessment into a speech recognition learning design to support students' English speaking skills, learning anxiety, and cognitive load. *Educational Technology & Society*, 25(1), 1–14. https://doi.org/10.30191/ETS.202201_25(1).0001

[5] Chiu, T. L., Liou, H. C., & Yeh, Y. (2007). A Study of web-based oral activities enhanced by Automatic Speech Recognition for EFL college learning. *Computer Assisted Language Learning*, 20(3), 209–233. https://doi.org/10.1080/09588220701489374

[6] Cohen, A., Oxford, R. L., & Chi, J. C. (2001). *Learning style survey: Assessing your own learning styles*. Center for Advanced Research on Language Acquisition, University of Minnesota.

[7] Dandu, G., Charyulu, G., Kumari, K., Dandu, G., Charyulu, G. M., & Kumari, K. L. (2024). AI-driven language learning: The impact of Rosetta Stone on ESL students' speaking proficiency and self-control. *Rupkatha Journal*, 16(4). https://doi.org/10.21659/rupkatha.v16n4.09

[8] de Vries, B. P., Cucchiarini, C., Bodnar, S., Strik, H., & van Hout, R. (2015). Spoken grammar practice and feedback in an ASR-based CALL system. *Computer Assisted Language Learning*, 28(6), 550–576. https://doi.org/10.1080/09588221.2014.889713

[9] Elimat, A. K., & AbuSeileek, A. F. (2014). Automatic speech recognition technology as an effective means for teaching pronunciation. *JALT CALL Journal*, 10(1), 21–47. https://doi.org/10.29140/jaltcall.v10n1.166

[10] Evers, K., & Chen, S. (2021). Effects of automatic speech recognition software on pronunciation for adults with different learning styles. *Journal of Educational Computing Research*, 59(4), 669–685. https://doi.org/10.1177/0735633120972011

[11] Griffiths, C., & Soruç, A. (2020). *Individual differences in language learning: A complex systems theory perspective*. Palgrave Macmillan. https://doi.org/10.1007/978-3-030-52900-0

[12] Hamuddin, B., & Dahler. (2018). Blogs as Powerful Learning Tools: The Perception from EFL Students in Riau Main Island Indonesia. *IOP Conference Series: Earth and Environmental Science*, 156(1), 12060. https://doi.org/10.1088/1755-1315/156/1/012060

[13] Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., ... & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6), 82–97. https://doi.org/10.1109/MSP.2012.2205597

[14] Jiang, M. Y. C., Jong, M. S. Y., Lau, W. W. F., Chai, C. S., & Wu, N. (2021). Using automatic speech recognition technology to enhance EFL learners' oral language complexity in a flipped classroom. *Australasian journal of educational technology*, 37(2), 110–131. https://doi.org/10.14742/ajet.6798

[15] Liu, X., Xu, M., Li, M., Han, M., Chen, Z., Mo, Y., Chen, X., & Liu, M. (2019). Improving English pronunciation via automatic speech recognition technology. *International Journal of Innovation and Learning*, 25(2), 126–140. https://doi.org/10.1504/IJIL.2019.097674

[16] McCrocklin, S. (2016). Pronunciation learner autonomy: The potential of automatic speech recognition. *System*, 57, 25–42. https://doi.org/10.1016/j.system.2015.12.013

[17] McCrocklin, S. (2019). Learners' feedback regarding ASR-based dictation practice for pronunciation learning. *CALICO Journal*, 36(2), 119–137. https://doi.org/10.1558/cj.34738

[18] Michael, C. (2017). Automated Speech Recognition in language learning: Potential models, benefits and impact[J]. *Training, Language and Culture*, 1(1), 46–61.

[19] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector

space. *arXiv*.Org. https://doi.org/10.48550/arxiv.1301.3781

[20] Nguyen, T.-H., Hwang, W.-Y., Pham, X.-L., & Ma, Z.-H. (2018). User-oriented EFL speaking through application and exercise: Instant speech translation and shadowing in authentic context. *Educational Technology & Society*, 21(4), 129–142.

[21] Rawas, S. (2024). AI: the future of humanity. *Discover Artificial Intelligence*, 4(1), 25–14. https://doi.org/10.1007/s44163-024-00118-3

[22] Shadiev, R., Wu, T.-T., Sun, A., & Huang, Y.-M. (2018). Applications of speech-to-text recognition and computer-aided translation for facilitating cross-cultural learning through a learning activity: Issues and their solutions. *Educational Technology Research and Development*, 66(1), 191–214. https://doi.org/10.1007/s11423-017-9556-8

[23] Tsai, S.-C. (2023). Learning with mobile augmented reality- and automatic speech recognition-based materials for English listening and speaking skills: Effectiveness and perceptions of non-English major English as a foreign language students. *Journal of Educational Computing Research*, 61(2), 444–465. https://doi.org/10.1177/07356331221111203

[24] Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478. https://doi.org/10.2307/30036540

[25] Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, 36(1), 157–178. https://doi.org/10.2307/41410412

[26] Yaniafari, R. P., Olivia, V., & S. (2022). The potential of ASR for improving English pronunciation: A review. *KnE Social Sciences*, 7(7), 281–289. https://doi.org/10.18502/kss.v7i7.10670

[27] Young, S., Evermann, G., Kershaw, D., Gales, M., Odell, J., Ollason, D., ... & Woodland, P. (2002). *The HTK book* (for HTK version 3.2). Cambridge University Engineering Department.

Author Profile

Yuan Fan (1989–), female, from Jiujiang, Jiangxi, China, is a lecturer with a master's degree, primarily engaged in foreign language teaching research.