

How Does ASR Technology Enhance English Learning Outcomes? Examining the Roles of Learner's Origin, Device Type, and Learning Styles

Yuan Fan

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

Abstract: *This study explores the impact of Automatic Speech Recognition (ASR) technology on English language learning among non-native speakers. We specifically investigate the roles of learners' geographic and cultural backgrounds, device type (high-end Apple and Android devices), and learning styles (visual, auditory) in mediating the efficacy of ASR. Employing a mixed-methods approach, our research integrates traditional teaching methods with ASR's real-time feedback to assess its effectiveness across different learner profiles in a controlled environment. Results indicate that while ASR technology significantly enhances linguistic accuracy and fluency, its success varies significantly depending on the learner's origin, the sophistication of the technology used, and the learner's preferred sensory modalities. Particularly, visual learners show the most substantial gains, suggesting that ASR technology integration should be tailored to individual learning styles to maximize educational outcomes. This study highlights the critical need for adaptive learning environments that align with diverse learner needs, providing key insights for educators and developers aiming to optimize technology-enhanced language learning platforms.*

Keywords: Automatic Speech Recognition (ASR) Technology, English Learning.

1. Introduction

1.1 Background

Spoken English has increasingly become a focal point in global education, transcending its role as merely a communication tool to become a key player in the international arena. With the number of second language English speakers surpassing that of native speakers, English has been adopted as an official language in 55 countries and is spoken in 75 countries worldwide. It is a mandatory subject in the school curricula of 138 countries and an optional subject in another 40⁺. Despite this widespread adoption, non-native speakers often face significant challenges due to the lack of interaction with native speakers and disparate learning environments, which can hinder their practical language use.

The advent of Automatic Speech Recognition (ASR) technology offers a promising solution to these challenges. Integrated into everyday smart devices, ASR technology provides real-time feedback and voice assessment, making it a potent tool for enhancing spoken English skills. Its effectiveness, however, varies among learners and lacks a standard evaluation criterion, prompting this study's exploration into its educational impacts. Research indicates that ASR technology, combined with peer correction, surpasses traditional teacher-led methods in teaching pronunciation and speaking skills in a second language (L2) context [1]. Moreover, its application in flipped classroom settings for vocabulary learning has shown to provide continuous feedback and reduce language learning anxiety, thereby fostering a more interactive and supportive learning environment [2].

This study aims to assess how various factors such as learners' backgrounds, technological devices, and learning styles influence the effectiveness of ASR technology in teaching spoken English. It explores whether students from metropolitan areas or those with international exposure are

better equipped to leverage ASR technology due to their broader language experience. Furthermore, the study examines the impact of device type - specifically high-end Apple or Android devices - and learning preferences (visual and auditory) on learners' English proficiency enhancements.

Ultimately, this study seeks to elucidate the moderating role of motivation in the relationship between learning effectiveness and ASR technology usage. By doing so, it aspires to provide novel insights into the application of educational technology and to establish an empirical foundation for teaching English as a second language, thereby guiding practical applications and helping learners more adeptly utilize ASR technology to improve their English speaking skills.

2. Theoretical Background of ASR and Education

ASR technology, which utilizes advanced computer algorithms including machine learning and deep learning, can recognize and transcribe human speech into text [3,4]. This capability allows computers to understand and respond to spoken inputs, thereby facilitating real-time feedback and evaluations. In educational settings, especially for ESL (English as a Second Language) and EFL (English as a Foreign Language) learners, ASR provides a valuable tool for practicing and improving pronunciation, intonation, and fluency. The technology's ability to offer immediate corrections on pronunciation errors enhances learners' language comprehension and speaking abilities [5, 6].

The integration of ASR technology has transformed traditional English teaching methodologies by introducing adaptive learning features that tailor instructions based on the learner's progress and capabilities. This adaptation is crucial in environments where interaction with native speakers is limited, thus providing a simulated and interactive learning

experience. Studies such as those by Liu (2021) have extended the use of ASR technology to specialized fields, demonstrating its efficacy in professional English contexts through tools like the “Naval Professional English Automatic Speech Evaluation System” [7]. Further, the application of ASR in interdisciplinary education has been explored, with technologies developed to convert spoken English into programming code, highlighting the versatile potential of ASR beyond conventional language learning [8].

Empirical research underscores the positive reception of ASR technology among students. For instance, Liu J, Liu X, Yang C (2022) documented students’ favorable perceptions towards using ASR for enhancing English speaking skills, noting its effectiveness in supporting their learning processes [9]. Additionally, the comparison between ASR systems and native listeners in evaluating the comprehensibility of L2 pronunciation reveals significant correlations, suggesting that ASR can reliably mimic human evaluative capabilities, which is crucial for validating ASR’s practical application in educational contexts [10].

3. Research Hypothesis

3.1 Learning Efficiency and Spoken English

Research across various studies underscores the significance of teacher-student interactions and the use of digital tools in enhancing English learning efficiency. For instance, applications such as corpus-based software in medical English vocabulary instruction and computer-assisted interactive methods in college English courses have been shown to significantly improve vocabulary acquisition and teaching efficiency [11, 12]. Moreover, evaluations of teacher effectiveness in Academic English Programs (EAP) have revealed a positive correlation between skilled teaching and student outcomes in English proficiency [13]. This body of evidence supports the hypothesis that learning efficiency positively correlates with improvements in spoken English skills, suggesting that enhanced educational strategies and teacher performance directly influence students’ ability to speak English proficiently [14].

H1: Learning efficiency positively correlates with spoken English.

3.2 ASR and Learning Efficiency

Recent advancements in ASR technology have significantly enhanced English learning, particularly in improving oral proficiency. Studies such as Jiang et al. (2022) demonstrate the benefits of ASR in flipped classrooms, where it has been shown to significantly enhance vocabulary complexity and fluency among Chinese university students [15]. Additionally, ASR’s impact on syntactic complexity was confirmed by Jiang et al. (2021), highlighting superior performance in EFL learners who utilized ASR technology compared to those who did not [16]. Surveys from Liu J, Liu X, and Yang C (2022) also reflect positive student perceptions regarding ASR, emphasizing its adaptability to diverse learning needs and its effectiveness in boosting learning efficiency [17]. Further research by Hannah et al. (2022) indicates that ASR accuracy varies significantly with the type of oral tasks and the learners’

language backgrounds, suggesting a nuanced application of ASR technology in language teaching [18].

H2: The use of ASR technology positively correlates with learning efficiency.

3.3 Learning Motivation

Recent studies highlight the critical role of learning motivation in enhancing English speaking skills through the use of ASR technology. Research by Lee and Drajeri (2019) underscored the influence of emotional factors, like motivation and confidence, on second language communication through informal digital English learning activities [19]. Moreover, Tejedor García et al. (2020) found that motivational elements in gamified ASR pronunciation training significantly improved learners’ pronunciation skills [20]. Further evidence from Evers and Chen (2020) suggests that learning styles impact the effectiveness of ASR, with visual learners outperforming verbal learners in tasks enhanced by ASR technology [21]. These findings support the hypothesis that learning motivation not only boosts learning efficiency but also modulates the efficacy of ASR technology in improving oral proficiency across diverse learner backgrounds. This synthesis of empirical data forms the foundation of our research framework, positioning learning motivation as a pivotal moderator in the relationship between ASR usage and English speaking improvement.

H3a: Learning motivation moderates the relationship between learning efficiency and spoken English.

H3b: Learning motivation moderates the relationship between ASR technology and learning efficiency.

3.4 Learner’s Origin and Spoken English

Research indicates that learners’ geographical and cultural exposure significantly impacts their English speaking skills. Studies by Stufft and Brogadir (2019) and Zhang and Lien (2020) demonstrate that residents of large cities and frequent travelers, who are often immersed in English-rich environments due to abundant educational resources and international exchanges, exhibit superior English oral proficiency [22, 23]. This advantage is further enhanced among those working in the highly internationalized tourism industry, where daily interactions with English speakers improve both the listening and speaking abilities of employees [24]. Based on this evidence, it is hypothesized that such individuals are likely to benefit more from ASR technology in improving their English speaking skills, due to their greater familiarity and confidence in using the language.

H4a: Students from first tier cities will perform better in using ASR technology to improve their spoken English.

H4b: Students who frequently travel abroad will use ASR technology to improve their spoken English more effectively.

3.5 Mobile Equipment and Spoken English

Recent studies highlight the significant role of device type and configuration in leveraging ASR technology to enhance

English speaking skills. Research from Lee (2018) and Klimova (2018) shows that students increasingly rely on high-end smartphones, such as Apple and Android devices, for learning English, particularly for vocabulary acquisition, due to the flexibility and personalized learning experiences these devices offer [25, 26]. Further investigations by Sad (2022) and Ayuningtyas (2018) emphasize that while smartphones boost learning motivation and engagement, dependency on these devices can pose challenges [27, 28]. However, the overall effectiveness of mobile applications in promoting interactive and collaborative learning supports the hypothesis that sophisticated mobile devices enhance the efficacy of ASR technology in improving spoken English proficiency among learners. Moreover, high-end devices: Devices with top-tier processors (e.g., Snapdragon 8 series, Apple A16 Bionic), 8GB RAM, and priced above \$800. These devices typically support advanced ASR app features, such as real-time transcription and high-accuracy speech recognition. Mid-end devices: Devices with mid-range processors (e.g., Snapdragon 7 series, Apple A14 Bionic), 4-8GB RAM, and priced between 400-800. These devices balance performance and affordability, supporting most ASR app functionalities. Low-end devices: Devices with entry-level processors (e.g., MediaTek Helio series), 4GB RAM, and priced below \$400. This classification is now explicitly stated in the revised manuscript. These devices may experience limitations in running resource-intensive ASR apps.

H5a: Students who use Apple's top of the line devices will have better performance in improving their spoken English using ASR technology compared to users of mid-range or low-end devices.

H5b: Students who use high-end Android devices have a more significant improvement in spoken English when using ASR technology compared to users of mid-range or low-end devices.

3.6 Learning Style and Spoken English

Research by Bakri et al. (2019), Masuram and Sripada (2020), and Ziadat (2021) has significantly advanced our understanding of how different learning styles- visual, auditory, kinesthetic, and tactile (VAKT) - affect English language proficiency [29-31]. These studies demonstrate that employing VAK and VAKT learning strategies in training can lead to marked improvements in students' English speaking accuracy, fluency, and comprehensibility, as well as in reading comprehension and oral reading skills. The effectiveness of these strategies highlights the crucial role of adapting teaching methods to match learning preferences, which in turn significantly enhances English speaking abilities. Based on this empirical evidence, it is hypothesized that specific learning styles can effectively improve spoken English skills among learners.

H6a: Students who prefer visual learning will improve their spoken English faster when using ASR technology.

H6b: Students who prefer auditory learning will improve their spoken English faster when using ASR technology.

Research model is shown in Figure 1.

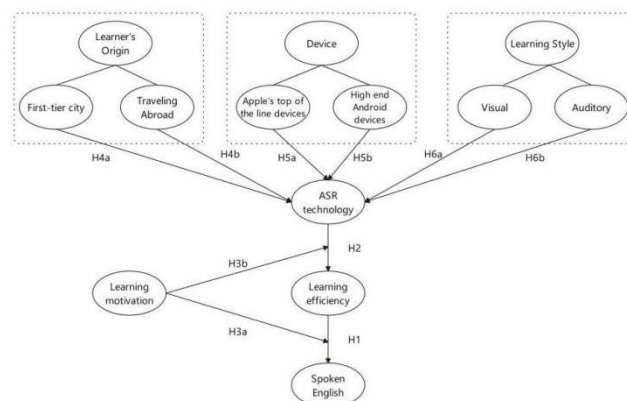


Figure 1: Research model

4. Experimental Results

This study involved multiple research categories, including metropolitan background, travel abroad experience, device type and usage frequency, device satisfaction, visual learning preferences, auditory learning preferences, use of ASR technology, learning efficiency, learning motivation, and improved English speaking ability. Each category was collected through a questionnaire survey with 315 samples, and the sources of the questionnaire items were adapted from previous studies as shown in Table 1, focusing on how to improve English speaking ability. The recruitment period for this study was from April 18, 2024, to May 20, 2024, at Lingnan Normal University, and the ethics committee waived the need for consent.

This study employed a combination of exploratory factor analysis (EFA), confirmatory factor analysis (CFA), and structural equation modeling (SEM) to investigate the factors influencing the improvement of English-speaking ability. First, potential influencing factors were identified through EFA. During the factor extraction process, principal component analysis (PCA) was applied, with an eigenvalue threshold greater than 1 set to ensure that each extracted factor significantly contributed to explaining the variance in the data. To enhance the interpretability of the factors, the Varimax rotation method was utilized, which maximizes the correlation between factors and their respective variables while minimizing cross-loadings. This step facilitated a clearer structure for subsequent analysis. All factor analyses were conducted using PSS software. Prior to extraction, the suitability of the data for factor analysis was confirmed through the Kaiser-Meyer-Olkin (KMO) test, which measures sampling adequacy, and Bartlett's test of sphericity, which assesses whether the correlation matrix is appropriate for factor analysis. Following these tests, factors were extracted and rotated to derive the final factor structure. Next, CFA was performed using Amos software to validate the factor structure identified in the EFA. A measurement model was constructed, and its fit to the data was evaluated using established goodness-of-fit indices, including the chi-square to degrees of freedom ratio (CMIN/DF), the root mean square error of approximation (RMSEA), and the comparative fit index (CFI). These indices ensured that the factor model adequately represented the observed data. Subsequently, SEM was applied to examine the causal relationships among key variables, such as ASR technology usage, learning efficiency, and learning motivation. SEM is particularly advantageous as

it allows for the simultaneous estimation of multiple causal pathways, providing a comprehensive understanding of how these variables collectively influence English-speaking ability. Model parameters were estimated using the maximum likelihood estimation (MLE) method, and path analysis was conducted in Amos software to ensure the model accurately reflected the hypothesized relationships. Significance testing of the path coefficients further elucidated the direct and indirect effects of ASR technology usage on learning efficiency and English-speaking ability. Through this rigorous statistical approach, the study demonstrates high credibility and scientific validity. The results provide robust evidence for the key factors affecting the improvement of English-speaking ability, offering valuable insights for both researchers and practitioners in the field of language learning.

4.1 Dataset Description

This study collected basic information about participants, including gender, age, educational background, English proficiency, frequency of learning tool usage, and device types, through questionnaire analysis. As shown in Figure 2, in terms of gender, 59% of the respondents were female, and 41% were male, reflecting a higher participation rate of women in educational technology research. Regarding age distribution, most respondents were young students aged 21 to 23 (39%), followed by those aged 18 to 20 (30%), 24 to 26 (27%), and 27 years or older (4%). In terms of educational background, undergraduates accounted for 59%, master’s students for 28%, and doctoral students for 13%, indicating that most participants were highly educated undergraduates. With respect to English proficiency, the majority of respondents were at an intermediate level 39%, followed by advanced 29% and beginner levels 30%, while only 2% were native or near-native speakers. Regarding the frequency of learning tool usage, 43% of respondents used the tools once a day, 29% twice, 16% three times, and 11% four times, suggesting that most respondents frequently utilized learning tools. In terms of device types, entry-level Apple devices accounted for 37%, high-end Apple devices for 30%, entry-level Android devices for 21%, high-end Android devices for 11%, and other devices for 1%. The prevalence of Apple devices among respondents may influence the applicability and effectiveness of ASR technology. The analysis of this basic information provides robust support for the subsequent interpretation of data.

Table 1: Measurement items and sources.

Constructs	Number of Measurement Items	Sources
Origin of Learners	6	S. Diana et al. 32, Gonzales et al. 33
Device Type and Usage Frequency	6	Meng Chen 34, Ayat Tarazi et al. 35, D. Burgos et al. 36
Learning Styles	6	Kazuya et al. 37, Yi-shan et al. 38, Setya et al. [39]
Use of ASR Technology	5	Jiaxin et al. [40], Xianxian et al. 41
Learning Efficiency	5	Xianxian Wu et al. [42], Jiaxin Liu et al. [43]
Learning Motivation	8	M. Ajmal [44], Sedina Selimovic [45] V. Halas et al. 46, S. Akmal et al. 47
Improved English Speaking Ability	5	M. Jong et al. 48, J. Beck et al. 49



Figure 2: Basic information of questionnaire

Table 2: Basic indicator - Obtaining the basic indicator by calculating the inverse scoring average across all dimensions.

Name	Sample Size	Minimum	Maximum	Average	Standard Deviation	Median
Metropolitan Background	315	1.000	5.000	2.277	1.051	2.000
Travel Abroad Experience	315	1.000	5.000	2.145	0.984	2.000
Device Type and Usage Frequency	315	1.000	5.000	2.329	1.047	2.000
Device Satisfaction	315	1.000	5.000	2.362	1.059	2.000
Visual Learning Preference	315	1.000	4.667	2.268	1.069	1.667
Auditory Learning Preference	315	1.000	4.667	2.125	0.976	1.667
Use of ASR Technology	315	1.000	4.600	2.317	1.017	1.800
Learning Efficiency	315	1.000	4.800	2.100	0.884	1.800
Learning Motivation	315	1.375	4.500	2.388	0.915	2.000
Improved English Speaking Ability	315	1.000	4.800	2.297	1.021	1.800

The statistical results from Table 2 show that the average values of various indicators are mostly concentrated between 2.000 and 2.400, indicating a relatively balanced distribution of the sample population across these metrics. The relatively small standard deviation values suggest a low degree of dispersion around the mean, indicating good consistency among the samples. The proximity of median values to the average further confirms the data's central tendency. Overall, this helps us to understand the overall characteristics and distribution of the sample population more comprehensively, presenting certain features and trends, and providing valuable references for further analysis of factors influencing the improvement of English speaking ability.

4.2 Factor Analysis

4.2.1 Reliability and validity analysis

Table 3: Reliability and validity results

Variables	Items	Alpha	KMO
Metropolitan Background	3	0.799	0.709
Travel Abroad Experience	3	0.781	0.702
Device Type and Usage Frequency	3	0.811	0.711
Device Satisfaction	3	0.789	0.703
Auditory Learning Preference	3	0.839	0.727
Use of ASR Technology	3	0.805	0.714
Auditory Learning Preference (Extended)	5	0.869	0.873
Learning Efficiency	5	0.851	0.869
Learning Motivation	8	0.906	0.938
Improved English-Speaking Ability	5	0.870	0.877
Total	41	0.916	0.888

Cronbach's alpha is used to assess the internal consistency of questionnaires or scales, with a coefficient value greater than 0.7 typically indicating good internal consistency. The Kaiser-Meyer-Olkin (KMO) test measures the suitability of data for factor analysis, where a KMO value closer to 1 suggests that the data is appropriate for factor analysis. Table 3 shows that the Cronbach's alpha values for all variables exceed 0.7, indicating that the questionnaire has good internal consistency and is suitable for subsequent analysis. Meanwhile, the KMO values are generally high, all exceeding 0.7, further confirming that the data is suitable for factor analysis. Specifically, the KMO values are as follows: metropolitan background (0.709), travel abroad experience (0.702), device type and usage frequency (0.711), device satisfaction (0.703), auditory learning preference (0.727), ASR technology use (0.714), visual learning preference (0.873), learning efficiency (0.869), learning motivation (0.938), and improvement in English speaking ability (0.877). All these KMO values demonstrate strong suitability of the data for factor analysis.

4.2.2 Analysis of factor numbers

In the process of factor analysis, as shown in Table 4 we successfully extracted 10 factors from the data, each with eigenvalues greater than 1, indicating that each factor can explain a certain amount of variability in the data. After rotating these factors, we obtained new ratios of each factor's explanation of data variance. The sum of squares load for rotated factors explained 68.644% of the variance, thereby effectively understanding the contribution of each factor to the overall data variability, and enhancing the explanatory power of the analysis.

Table 4: Factor number analysis results Extracted load sum of squares

Factor	Eigenvalue	Variance Explained %	Cumulative %
1	9.736	23.747	23.747
2	4.743	11.567	35.315
3	2.443	5.959	41.274
4	2.068	5.043	46.316
5	2.017	4.920	51.236
6	1.653	4.033	55.269
7	1.537	3.748	59.017
8	1.416	3.453	62.470
9	1.307	3.189	65.659
10	1.224	2.986	68.645
Rotated load sum of squares			
Factor	Eigenvalue	Variance Explained %	Cumulative %
1	4.948	12.069	8.403
2	3.445	8.403	20.472
3	3.343	8.153	28.610
4	3.336	8.137	36.763
5	2.263	5.520	42.229
6	2.241	5.466	47.666
7	2.229	5.437	53.081
8	2.220	5.415	58.124
9	2.068	5.043	63.645
10	2.050	5.000	68.644

4.3 Confirmatory Analysis

4.3.1 Model indicator analysis

This section conducts confirmatory factor analysis (CFA) on the variables to assess the model's goodness-of-fit. According to the results presented in Table 5, the CMIN/DF value is 1.172, which is significantly lower than the ideal threshold of 3, indicating a good fit of the model. CMIN/DF is a critical metric for evaluating model fit, and a value below 3 generally suggests a satisfactory alignment between the model and the data. Thus, the result of 1.172 demonstrates that our model meets the standard for goodness-of-fit. Furthermore, the RMSEA value is 0.023, which is well below the benchmark of 0.08, further confirming the model's excellent fit. Therefore, the RMSEA result of 0.023 indicates that the model's fit is highly ideal and fully meets the required standards.

Table 5: Model fit indicators and their judgment values

Model Fit	CMIN	DF	CMIN/DF	NFI	RFI	IFI	TLI	CFI	GFI	RMSEA
Fit Results	859.965	734	1.172	0.872	0.857	0.979	0.976	0.979	0.886	0.023
Judgment Values			<3	>0.9	>0.9	>0.9	>0.9	>0.9	>0.9	<0.08

However, although the model’s NFI, RFI, and GFI values are 0.872, 0.857, and 0.886, respectively, which are slightly below the ideal threshold of 0.9, these values are still considered acceptable. In practical data analysis, values close to 0.9 are often deemed satisfactory, particularly in cases with large sample sizes or complex data structures. Thus, while these indices do not fully reach 0.9, they still indicate a good model fit that meets the requirements for practical application. Additionally, the IFI, TLI, and CFI values are 0.979, 0.976, and 0.979, respectively, all exceeding 0.9, which signifies a very high level of model fit and excellent overall performance.

In sum, although the NFI, RFI, and GFI values are slightly lower, their proximity to 0.9, combined with the strong performance of other fit indices (such as IFI, TLI, and CFI), confirms that the model’s overall goodness-of-fit is acceptable. This supports the validity of our hypothesis testing and the examination of relationships among variables.

4.3.2 Convergent validity analysis

To evaluate the convergent validity of the variables in the model, this study conducted Average Variance Extracted (AVE) and Composite Reliability (CR) analyses. The factor loadings, which measure the representativeness and contribution of each item to the construct, were all greater than 0.7, indicating strong validity for each item. CR was used to assess the internal consistency of the items within each construct, with higher values indicating stronger reliability. In this study, the CR values for all constructs exceeded 0.7, demonstrating good internal consistency. As shown in Table 6, for example, the CR value for learning motivation was 0.906, significantly higher than the threshold of 0.7, indicating high measurement reliability. AVE was used to measure the convergent validity of the constructs, with higher values indicating that the construct effectively explains its underlying variables. In this study, the AVE values for most constructs exceeded 0.5, demonstrating good convergent validity and the ability to effectively explain the variance of related variables. For instance, the AVE for learning motivation was 0.548, indicating that it adequately captures the related factors.

Table 6: Convergent Validity Results

Construct	Item	Loading Factor	CR	AVE
Metropolitan Background	MB1	0.735	0.571	0.799
	MB2	0.740		
	MB3	0.790		
Travel Abroad Experience	TAE1	0.754	0.545	0.782
	TAE2	0.697		
	TAE3	0.761		
Device Type and Usage Frequency	DTUF1	0.784	0.591	0.812
	DTUF2	0.720		
	DTUF3	0.800		
Device Satisfaction	DS1	0.657	0.553	0.786
	DS2	0.734		
	DS3	0.829		

Auditory Learning Preference	ALP1	0.813	0.651	0.848
	ALP2	0.768		
	ALP3	0.837		
Use of ASR Technology	UASRT1	0.827	0.602	0.819
	UASRT2	0.745		
	UASRT3	0.752		
Auditory Learning Preference	ALP1	0.810	0.572	0.870
	ALP2	0.744		
	ALP3	0.720		
	ALP4	0.731		
	ALP5	0.772		
Learning Efficiency	LE1	0.793	0.545	0.857
	LE2	0.697		
	LE3	0.782		
	LE4	0.707		
	LE5	0.707		
Learning Motivation	LM1	0.726	0.548	0.906
	LM2	0.776		
	LM3	0.703		
	LM4	0.710		
	LM5	0.746		
	LM6	0.735		
	LM7	0.766		
	LM8	0.757		
Improved English Speaking Ability	IESA1	0.784	0.572	0.870
	IESA2	0.769		
	IESA3	0.762		
	IESA4	0.738		
	IESA5	0.728		

In the construct of travel abroad experience, the factor loadings for TAE1 and TAE3 were greater than 0.7, while the loading for TAE2 was 0.697. Despite this, the overall construct still exhibited strong internal consistency, with a CR of 0.782 and an AVE of 0.545, confirming its reliability and effectiveness in explaining the underlying factors. Similarly, the constructs of device type and usage frequency met the required standards for factor loadings, CR, and AVE, further validating the efficiency and effectiveness of the questionnaire. Overall, the questionnaire demonstrated strong internal consistency and convergent validity, effectively measuring each construct, and ensuring the reliability and validity of the measurement results.

4.3.3 Discriminant validity analysis

By comparing the Pearson correlations between variables with the square root of their AVE, we can assess the discriminant validity among variables. Ideally, the square root of an latent variable’s AVE should be greater than its correlations with any other variable, indicating good discriminant validity.

As shown in Table 7, the square root values of AVE for all variables (bold on the diagonal) are greater than the Pearson correlation coefficients between the respective variable and all other variables, indicating good discriminant validity among the variables in this study, i.e., each latent variable is distinct compared to other variables.

Table 7: Pearson correlation and square root of AVE values

	1	2	3	4	5	6	7	8	9	10
1 Metropolitan Background	0.756									
2 Travel Abroad Experience	0.337	0.738								
3 Device Type and Usage Frequency	0.299	0.423	0.769							
4 Device Satisfaction	0.436	0.325	0.414	0.744						
5 Visual Learning Preference	0.383	0.488	0.413	0.433	0.807					
6 Auditory Learning Preference	0.284	0.178	0.231	0.312	0.277	0.776				
7 ASR Technology Use	0.393	0.449	0.445	0.399	0.515	0.279	0.756			
8 Learning Efficiency	0.306	0.457	0.465	0.389	0.432	0.185	0.387	0.738		
9 Learning Motivation	0.251	0.028	0.054	0.237	0.032	0.095	0.292	0.267	0.740	
10 Improvement in English Speaking Ability	0.458	0.576	0.349	0.390	0.478	0.315	0.454	0.392	0.076	0.756

4.4 Structural Model

4.4.1 Structural equation model path analysis

This study further analyzed the path coefficients of the structural equation model. First, after confirming the good fit of the model, the significance of the paths was evaluated using indicators such as unstandardized path coefficients, standard errors (S.E.), critical ratios (C.R.), p-values, and standardized path coefficients. As shown in Table 8 the impact of overseas experience on ASR technology use was significant, with a path coefficient of 0.206, a C.R. of 2.448, and a p-value of 0.014, supporting the hypothesis. The influence of device type and frequency of use was also significant, with a path

coefficient of 0.194, a C.R. of 2.789, and a p-value of 0.005, further supporting the hypothesis. The effect of visual learning preference on ASR technology use was highly significant, with a path coefficient of 0.228, a C.R. of 3.523, and a p-value of 0.000, strongly supporting the hypothesis. In contrast, metropolitan background, device satisfaction, and auditory learning preference did not show significant effects on ASR technology use. The path coefficient for metropolitan background was 0.133, with a C.R. of 1.879 and a p-value of 0.06, failing to reach significance. The path coefficients for device satisfaction and auditory learning preference were 0.087 and 0.081, respectively, with C.R. values of 1.237 and 1.225 and p-values greater than 0.05, indicating no significant impact on ASR technology use.

Table 8: Structural equation model path coefficients test results

Path	Unstandardized Path Coefficient	S.E.	C.R.	P	Standardized Path Coefficient	Conclusion
ASR Technology Use ← Metropolitan Background	0.133	0.071	1.879	0.06	0.128	Not Supported
ASR Technology Use ← Travel Abroad Experience	0.206	0.084	2.448	0.014	0.178	Supported
ASR Technology Use ← Device Type and Usage Frequency	0.194	0.070	2.789	0.005	0.193	Supported
ASR Technology Use ← Device Satisfaction	0.087	0.070	1.237	0.216	0.089	Not Supported
ASR Technology Use ← Visual Learning Preference	0.228	0.065	3.523	0.000***	0.257	Supported
ASR Technology Use ← Auditory Learning Preference	0.081	0.066	1.225	0.221	0.074	Not Supported
Learning Efficiency ← ASR Technology Use	0.316	0.045	6.947	0.000***	0.454	Supported
Improvement in English Speaking Ability ← Learning Efficiency	0.593	0.094	6.32	0.000***	0.419	Supported

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$. Use a 95% confidence level ($\alpha = 0.05$).

Further analysis revealed that ASR technology use significantly improved learning efficiency, with a path coefficient of 0.316, a C.R. of 6.947, and a p-value of 0.000, demonstrating a significant positive effect of ASR technology on learning efficiency. Learning efficiency, in turn, had a significant impact on the improvement of English-speaking ability, with a path coefficient of 0.593, a C.R. of 6.32, and a p-value of 0.000, highlighting the critical role of learning efficiency in enhancing oral proficiency. Overall, significant relationships were found between overseas experience, device type and frequency of use, visual learning preference, and ASR technology use, while metropolitan background, device satisfaction, and auditory learning preference did not significantly influence ASR technology use. ASR technology significantly enhanced learning efficiency, which subsequently contributed to the improvement of English-speaking ability.

4.4.2 The mediating role of learning motivation in the effect of ASR technology use on learning efficiency

This study examines the mediating role of learning motivation

in the impact of ASR technology use on learning efficiency. Table 9 presents the results of three models. Model 1 validates the direct effect of ASR technology use on learning efficiency, with a path coefficient of 0.296, a C.R. of 6.404, and a p-value of less than 0.01, indicating a significant positive influence of ASR technology use on learning efficiency. However, this model does not account for learning motivation as a mediating variable. In Model 2, learning motivation is introduced as a moderating variable, revealing a path coefficient of 0.142, a C.R. of 2.695, and a p-value of less than 0.01, suggesting that learning motivation significantly and positively affects learning efficiency. While learning motivation plays a partial role in the relationship between ASR technology use and learning efficiency, it does not fully explain the impact of ASR technology use on learning efficiency. Model 3 further incorporates the interaction term between ASR technology use and learning motivation, showing a path coefficient of 0.143, a C.R. of 2.783, and a p-value of less than 0.01. This indicates that learning motivation significantly moderates the effect of ASR technology, implying that students with higher motivation may achieve greater improvements in learning efficiency through ASR technology.

Table 9: The mediating effect of learning motivation on the impact of ASR technology use on learning efficiency

	Model 1	Model 2	Model 3
Constant	2.100**(44.781)	2.100** (45.227)	2.065** (43.356)
ASR Technology Use	0.296** (6.404)	0.262** (5.530)	0.238** (5.003)
Learning Motivation		0.142** (2.695)	0.108** (2.028)
ASR Technology Use * Learning Motivation			0.143** (2.783)
Sample Size	315	315	315
R ²	0.116	0.136	0.157
Adjusted R ²	0.113	0.130	0.149
F Value	F (1,313)=41.014,p<0.000	F (2,312)=24.549,p<0.000	F (3,311)=19.302,p<0.000
ΔR ²	0.116	0.020	0.021
ΔF Value	F (1,313)=41.014,p<0.000	F (1,312)=7.263,p=0.007	F (1,311)=7.746,p=0.006

* $p < 0.05$; ** $p < 0.01$. Values in parentheses are t values. Use a 95% confidence level $\alpha = 0.05$.

The model fit was also assessed. Model 1 demonstrates an R² of 0.116, an adjusted R² of 0.113, an F-value of 41.014, and a p-value of less than 0.000, indicating a good model fit. For Model 2 and Model 3, the R² values are 0.136 and 0.157, respectively, suggesting that the inclusion of learning motivation as a moderating variable enhances the explanatory power of the model. The addition of the interaction term further improves the model's ability to explain variations in learning efficiency. Both F-values and p-values are less than 0.000, indicating significant model improvements. In sum, learning motivation plays a significant moderating role in the relationship between ASR technology use and learning efficiency. The model improvements enhance the explanatory power regarding learning efficiency, and the results support the notion that learning motivation is a critical factor influencing the effectiveness of ASR technology use.

4.4.3 The mediating role of learning motivation in the effect of learning efficiency on improvement of english speaking ability

This study examines the moderating role of learning motivation in the relationship between learning efficiency and the improvement of English-speaking ability. As shown in Table 10, Model 1 confirms a significant positive effect of learning efficiency on the improvement of English-speaking ability, with a path coefficient of 0.400, a C.R. value of 6.526, and a p-value of less than 0.01, indicating a significant positive correlation between learning efficiency and the enhancement of English-speaking ability. In Model 2, learning motivation is introduced as a moderating variable. The results reveal that learning motivation itself does not have a significant direct effect on the improvement of English-speaking ability, with a path coefficient of -0.012, a

C.R. value of -0.193, and a p-value of 0.847, suggesting that learning motivation does not directly influence the enhancement of English-speaking ability. Model 3 further incorporates the interaction term between learning efficiency and learning motivation. The results demonstrate that the interaction term has a path coefficient of 0.180, a C.R. value of 2.703, and a p-value of less than 0.01, indicating that learning motivation significantly moderates the effect of learning efficiency on the improvement of English-speaking ability. Students with higher levels of learning motivation can benefit more effectively from improvements in learning efficiency, thereby enhancing their English-speaking ability. These findings suggest that learning motivation is not only significantly correlated with learning efficiency but also plays a crucial moderating role in the impact of learning efficiency on the improvement of English-speaking ability.

Additionally, the R² value for Model 1 is 0.120, with an adjusted R² of 0.117, indicating that Model 1 explains approximately 12% of the variance. The R² values for Model 2 and Model 3 are 0.120 and 0.140, respectively, demonstrating that the inclusion of the moderating effect of learning motivation improves the explanatory power of the model. The F-values for the models are 42.587, 21.247, and 16.886, respectively, with p-values all below 0.000, indicating that the models are well-fitted, and the results are statistically significant. The increase in R² from 0.120 to 0.140 further validates the improvement of the model after incorporating the moderating effect of learning motivation. In conclusion, learning motivation significantly moderates the effect of learning efficiency on the improvement of English-speaking ability, particularly among students with higher levels of learning motivation, where the impact of learning efficiency on speaking ability enhancement is more pronounced.

Table 10: The mediating effect of learning motivation on the impact of learning efficiency on improvement of english speaking ability

	Model 1	Model 2	Model 3
Constant	2.297** (42.500)	2.297** (42.435)	2.264** (41.208)
Learning Efficiency	0.400** (6.526)	0.402** (6.390)	0.392** (6.269)
Learning Motivation		-0.012 (-0.193)	-0.061 (-0.975)
Learning Efficiency * Learning Motivation			0.180** (2.703)
Sample Size	315	315	315
R ²	0.120	0.120	0.140
Adjusted R ²	0.117	0.114	0.132
F Value	F (1,313)=42.587,p=0.000	F (2,312)=21.247,p=0.000	F (3,311)=16.886,p=0.000
ΔR ²	0.120	0.000	0.020
ΔF Value	F (1,313)=42.587,p=0.000	F (1,312)=0.037,p=0.847	F (1,311)=7.305,p=0.007

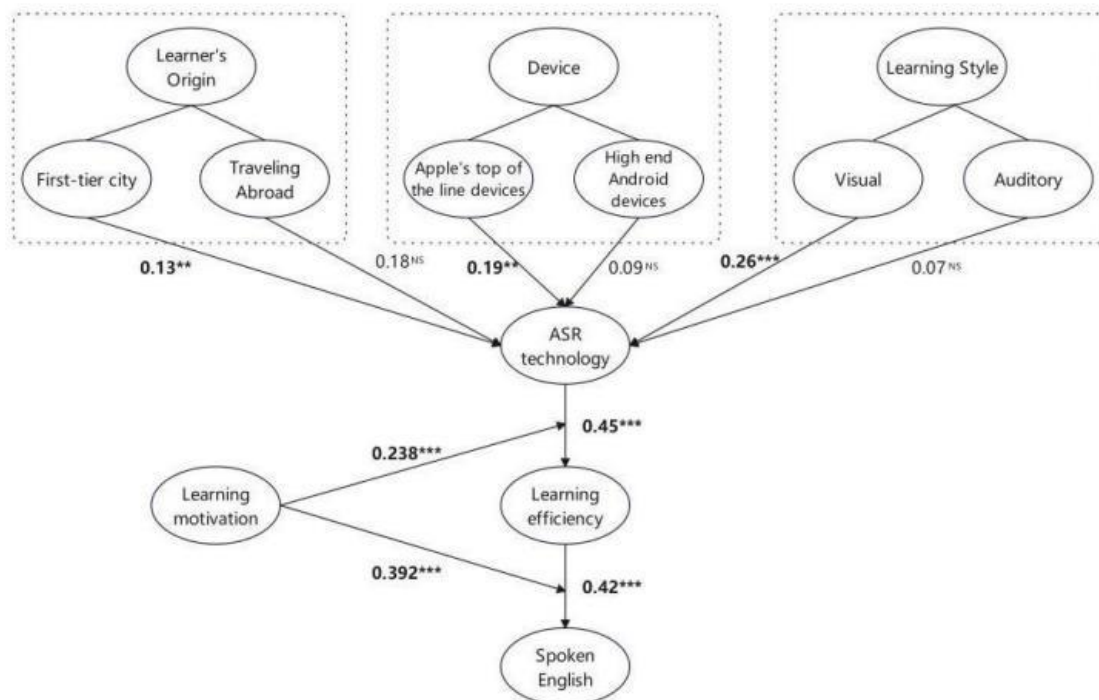


Figure 3: Hypothesis path coefficients

Table 11: Hypothesis testing results

Code	Hypothesis	Conclusion
H1	The learning efficiency of students is positively correlated with the improvement of English speaking ability.	Supported
H2	The use of ASR technology is positively correlated with the improvement of English speaking ability.	Supported
H3a	Learning motivation moderates the relationship between learning efficiency and improvement of English speaking ability.	Supported
H3b	Learning motivation moderates the relationship between ASR technology use and improvement of English speaking ability.	Supported
H4a	Students from metropolitan areas perform better in improving English speaking ability through the use of ASR technology.	Supported
H4b	Students with frequent travel abroad experiences achieve better outcomes in improving English speaking ability through ASR technology.	Not supported
H5a	Students using high-end Apple devices achieve better outcomes in improving English speaking ability through ASR technology compared to mid-range or low-end device users.	Supported
H5b	Students using high-end Android devices do not significantly improve their speaking ability through ASR technology compared to mid-range or low-end device users.	Not supported
H6a	Students with a preference for visual learning improve their English speaking ability faster through the use of ASR technology.	Supported
H6b	Students with a preference for auditory learning do not improve their English speaking ability faster through the use of ASR technology.	Not supported

4.5 Discussion

Through factor analysis, confirmatory analysis, moderation effect analysis, and path analysis of the structural equation model, this study has verified the hypothesis results as shown in Table 11 and Figure 3 path coefficients. It revealed complex relationships among different variables, especially the crucial roles of learning motivation and ASR technology use in improving English speaking ability, including:

- There is a significant positive correlation between learning efficiency and the improvement of English speaking ability, validating the importance of learning efficiency for language learning outcomes (H1). Therefore, teachers should design teaching strategies that can improve learning efficiency, promote the improvement of oral skills by strengthening feedback and stimulating learning motivation.
- The use of ASR technology has a significant positive impact on the improvement of English-speaking ability, supporting the application value of technology in language learning (H2). The path coefficient of the use

of ASR technology on learning efficiency is 0.316, with a P value less than 0.001, indicating that ASR technology also plays a significant role in improving learning efficiency, especially in helping students identify and correct pronunciation errors through immediate feedback. Although the path coefficient is lower than the impact of learning efficiency on oral ability, research has shown that ASR technology effectively improves learning efficiency through personalized feedback and interactivity, enhances students' sense of participation, and reduces anxiety in language learning 50.

- Learning motivation significantly moderates the relationship between learning efficiency and improvement of English-speaking ability (H3a), as well as between ASR technology use and improvement of English-speaking ability (H3b), highlighting the key role of motivation in the learning process. Learners with strong motivation typically exhibit higher levels of engagement, perseverance, and focus, which naturally contribute to improved learning outcomes, regardless of the specific tools or methods used, including ASR

techniques. This study confirms the moderating role of motivation in the learning context supported by ASR, revealing finer insights into the interaction between motivation and technology enhanced learning. This will help to go beyond general cognitive assumptions and gain a deeper understanding of how to optimize ASR techniques to support learners with different levels of motivation.

- Travel abroad experience and preference for visual learning have been proven to significantly enhance the effect of ASR technology use (H4b, H6a), while metropolitan background, device configuration level, and auditory learning preference did not reach a significant level of influence on ASR technology use (H4a, H5b, H6b). Although people may expect auditory learners to benefit greatly from ASR technology due to their reliance on auditory feedback, the results indicate the opposite. First, the ASR tool used in this study may not have fully utilized the advantages of auditory learners, such as their preferences for listening and speech repetition. Next, auditory learners may require more tailored feedback or interactive features, such as real-time pronunciation correction or adaptive difficulty levels, to fully benefit from ASR technology. Finally, the design of this study may not have captured the specific ways in which auditory learners use ASR tools, such as their dependence on auditory cues exceeding visual or textual support. Future research should explore how to customize ASR technology to better meet the learning preferences and needs of auditory learners, possibly through more immersive auditory experiences or personalized feedback mechanisms.

These conclusions provide valuable insights for English teaching practices, especially in the field of technology-assisted language learning. Future research could further explore how learning motivation can be enhanced through different educational intervention strategies, and how ASR technology applications can be optimized to better serve learners with different backgrounds and preferences.

5. Implications

5.1 Theoretical Implications

This study significantly advances our understanding of the effectiveness of ASR technology in English language learning, emphasizing the influence of learners' origins, device types, and learning styles. By integrating ASR technology, our findings demonstrate substantial improvements in pronunciation, fluency, and comprehension among non-native speakers, particularly when learning environments are tailored to individual learner profiles. The research shows that learners from diverse geographic backgrounds utilize ASR technology differently, which suggests that cultural and regional characteristics play critical roles in technology adoption and educational success. Similarly, the type of device used (high-end vs. mid-range) significantly affects learning outcomes, with high-end devices offering more robust support and better user experience. Moreover, the differential impact of ASR technology on various learning styles-particularly the pronounced benefits for visual

learners-highlights the need for adaptive educational technologies that cater to individual learning preferences. These findings not only underscore the multifaceted role of ASR technology in enhancing language learning but also align with and expand upon existing research, such as the work by Suess et al. 51, Gray et al. 52, and Mirheidari et al. [53], by offering a comprehensive analysis of how personalized ASR interventions can optimize learning outcomes. This comprehensive approach provides a clearer understanding of the cognitive and affective responses elicited by ASR technology, setting the stage for future research into personalized learning environments.

5.2 Practical Implications

Although high-end devices may provide clearer feedback and more interactive learning experiences, addressing the limitations of mid- and low-end devices is crucial to ensure fair access to ASR technology. Mid- and low-end devices often face challenges such as reduced processing power, decreased audio quality, and limited compatibility with advanced ASR features, which may hinder the effectiveness of ASR supported learning activities. For example, these devices may have difficulty performing real-time speech recognition or providing detailed feedback on pronunciation accuracy. Despite these limitations, educators can adopt strategies to alleviate these issues, such as optimizing ASR applications for low spec devices, integrating offline functionality, or designing activities that require minimal computing resources. In addition, educators should consider utilizing cloud-based ASR solutions that can offload processing requirements to remote servers, making it easier for users of mid- and low-end devices to access advanced features.

Moreover, owning high-end devices may reflect a better family financial situation and may provide additional English learning resources for children, such as extra tutoring. However, learning success depends not only on economic conditions or the technology used, but also on the complex interplay of personal effort, educational resources, teaching quality, and the adaptability of technology to different environments. Therefore, this study emphasizes the importance of addressing device related barriers and ensuring that ASR technology is accessible and effective for all learners, regardless of their socioeconomic status or device type. Thus, educators can create a more equitable and inclusive learning environment, maximizing the potential of ASR technology for different groups of learners. Overall, this study provides practical guidance for educators to design and implement ASR supported activities to adapt to learners' different technological and socio-economic realities. Future research should further explore strategies to optimize ASR technology for mid- and low-end devices, ensuring that everyone can benefit from ASR supported learning.

6. Limitations and Future Research Directions

While this study provides new insights into the application of ASR technology in the domain of English learning, it has several limitations that open directions for future research. Firstly, this study is primarily based on data from a single cultural background (China), which may limit the

generalizability of its conclusions. Future research could be conducted in different cultural and linguistic contexts, such as Southeast Asian countries, to verify and extend the findings of this study. Incorporating cross-cultural research would not only broaden the relevance of the findings but also provide a deeper understanding of how cultural and linguistic diversity influences the effectiveness of ASR technology in English learning. Second, the cross-sectional study design used in this study limits the inference of causal relationships. Although this design provides valuable insights into the relationship between ASR technology usage and English learning outcomes at specific time points, it fails to capture the dynamic and long-term impact of ASR technology on language learning. Future research should consider adopting a longitudinal study design to examine how the impact of ASR technology evolves over time. Longitudinal studies will enable researchers to track changes in learners' proficiency, engagement, and motivation, thereby gaining a more comprehensive understanding of the ongoing advantages or potential disadvantages of ASR technology. In addition, experimental methods such as randomized controlled trials can be used to establish causal relationships and isolate the specific effects of ASR technology from other confounding variables. Third, although this study explored the interactions between ASR technology, learner traits, and the learning environment, it did not cover all factors that might affect learning outcomes comprehensively. Hence, future research could explore more potential factors, such as individual learning strategies, the complexity of learning tasks, and the duration of technology use, on learning effectiveness. Furthermore, considering the rapid development of technology, future research should assess the potential and challenges of emerging ASR technologies and tools in teaching practices. As ASR systems become more advanced, incorporating features such as real-time feedback, adaptive learning algorithms, and multilingual support, it is crucial to evaluate their effectiveness in diverse educational settings. Such research could inform the development of more inclusive and accessible ASR tools that cater to the needs of learners from varied socioeconomic and cultural backgrounds.

7. Conclusion

This study explored the application of ASR technology in teaching spoken English and its impact on learning efficiency, proposing an integrated framework to analyze how ASR technology, learner backgrounds, device usage habits, and learning styles collectively contribute to improving spoken English abilities. By analyzing the effectiveness of ASR technology in different learning environments, this study revealed the potential of technology applications in language learning and their practical application value in educational practices. It further emphasized the critical role of learning motivation in the application of ASR technology, as well as the wide applicability and limitations of technology applications across cultural backgrounds.

Data Availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

References

- [1] Sun, W. The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: a mixed methods investigation. *Frontiers in Psychology*. 2023;14.
- [2] Jiang, M. Y. C., Jong, M. S. Y., Wu, N., et al. Integrating automatic speech recognition technology into vocabulary learning in a flipped English class for Chinese college students. *Frontiers in Psychology*. 2022; 13: 902429.
- [3] Kumar, Y., Singh, N. A comprehensive view of automatic speech recognition system-A systematic literature review. 2019 International Conference on Automation, Computational and Technology Management (ICACTM). IEEE. 2019:168-173.
- [4] Alharbi, S., Alrazgan, M., Alrashed, A., et al. Automatic speech recognition: Systematic literature review. *IEEE Access*. 2021; 9:131858-131876.
- [5] Liu, J., Liu, X., Yang, C. A study of college students' perceptions of utilizing automatic speech recognition technology to assist English oral proficiency. *Frontiers in Psychology*. 2022; 13: 1049139. *Frontiers*.
- [6] Kochem, T., Beck, J., Goodale, E. The Use of ASR-Equipped Software in the Teaching of Suprasegmental Features of Pronunciation: A Critical Review. *CALICO Journal*. 2022;39(3).
- [7] Liu, Haiqi. Design of Automatic Speech Evaluation System of Professional English for the Navy based on Intelligent Recognition Technology. 2021 2nd International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI). 2021:236-241.
- [8] Nizzad, A., Thelijjagoda, S. Designing of a Voice-Based Programming IDE for Source Code Generation: A Machine Learning Approach. 2022 International Research Conference on Smart Computing and Systems Engineering (SCSE). 2022:14-21.
- [9] Liu, J., Liu, X., Yang, C. A study of college students' perceptions of utilizing automatic speech recognition technology to assist English oral proficiency. *Frontiers in Psychology*. 2022; 13: 1049139. *Frontiers*.
- [10] Inceoglu, S., Chen, W-H., Lim, H. Assessment of L2 intelligibility: Comparing L1 listeners and automatic speech recognition. *ReCALL*. 2022; 35: 89-104.
- [11] Liu, C. Approach of the corpus research on medical English prepositions. 2019 International Conference on Arts, Management, Education, and Innovation, (ICAMEI 2019). 2019:1187-1191.
- [12] Fang, Y. Analysis on Teacher-student Interaction with Computer Aid in College English Language Teaching Class. *Journal of Physics: Conference Series*. 2021; 1744(4): 042089. IOP Publishing.
- [13] Ghaffarian Asl, S., Osam, N. A Study of Teacher Performance in English for Academic Purposes Course: Evaluating Efficiency. *SAGE Open*. 2021; 11(4): 21582440211050386. SAGE Publications.
- [14] Ming-mei, W. The Effect of Language Proficiency on the Teaching Knowledge Acquisition of English Student Teachers-Comparison of Three Repetitive Lessons. *Journal of Literature and Art Studies*. 2019; 9(4): 404-411.
- [15] Jiang, M. Y. C., Jong, M. S. Y., Wu, N., et al. Integrating automatic speech recognition technology into

- vocabulary learning in a flipped English class for Chinese college students. *Frontiers in Psychology*. 2022; 13: 902429. Frontiers.
- [16] Jiang, M. Y. C., Jong, M. S. Y., Lau, W. W. F., et al. Using automatic speech recognition technology to enhance EFL learners' oral language complexity in a flipped classroom. *Australasian Journal of Educational Technology*. 2021;37(2):110-131.
- [17] Liu, J., Liu, X., Yang, C. A study of college students' perceptions of utilizing automatic speech recognition technology to assist English oral proficiency. *Frontiers in Psychology*. 2022;13:1049139. Frontiers.
- [18] Hannah, L., Kim, H., Jang, E. E. Investigating the effects of task type and linguistic background on accuracy in automated speech recognition systems: Implications for use in language assessment of young learners. *Language Assessment Quarterly*. 2022;19(3):289-313. Taylor & Francis.
- [19] Lee, J. S., Drajiati, N. A. Affective variables and informal digital learning of English: Keys to willingness to communicate in a second language. *Australasian Journal of Educational Technology*. 2019;35(5):168-182
- [20] Gonzales, W., Hughes, M. T. Leveraging a Spanish Literacy Intervention to Support Outcomes of English Learners. *Reading Psychology*. 2021; 42: 411-434. doi:10.1080/02702711.2021.1888352.
- [21] Evers, K., Chen, S. Effects of automatic speech recognition software on pronunciation for adults with different learning styles. *Journal of Educational Computing Research*. 2021;59(4):669-685. SAGE Publications..
- [22] Stufft, D. L., Brogadir, R. Education the Culturally and Linguistically Diverse Non-Urban Population: Three Cost-Effective Strategies. *The Rural Educator*. 2010; 31(2).
- [23] Zhang, W., Lien, D. English listening, speaking, and earnings among workers in urban China. *Education Economics*. 2020;28(2):211-223. Taylor & Francis.
- [24] Bachore, M. M. English language literacy skills and academic achievement of urban and rural secondary schools: the case of high and low achievers. *Education Research International*. 2022;2022.
- [25] Lee, S., Koo, Y. College EFL students' perceptions of vocabulary learning with a smartphone application. *The Journal of Studies in Language*. 2018;34(3):377-392.
- [26] Klimova, B. Mobile phones and/or smartphones and their apps for teaching English as a foreign language. *Education and Information Technologies*. 2018; 23: 1091-1099. Springer.
- [27] Sad, S. N., Özer, N., Yakar, Ü., et al. Mobile or Hostile? Using Smartphones in Learning English as a Foreign Language. *Computer Assisted Language Learning*. 2022; 35: 1031-1057. Taylor & Francis.
- [28] Ayuningtyas, P. Whatsapp: Learning on the go. *Metathesis: journal of English language, literature, and teaching*. 2018;2(2):159-170.
- [29] Bakri, R. A., Rahman, M. A., Jabu, B. Exploring the Impact of VAK Learning Style on Teenager Level Language Learners in Indonesia. *Journal of Language Teaching and Research*. 2019;10(4):807-814.
- [30] Masuram, J., Sripada, P. N. Developing speaking skills through task-based materials. *Procedia Computer Science*. 2020;172:60-65. Elsevier.
- [31] Ziadat, A. H. The impact of using VAKT strategy on oral reading and reading comprehension skills of elementary students with dyslexia. *International Journal of Learning, Teaching and Educational Research*. 2021; 20(2):121-136.
- [32] Hidayati, T., Diana, S., Husna, F., Perrodin, D. D. Factors affecting English performance between students residing in tourist and non-tourist areas. *Studies in English Language and Education*. 2023. doi:10.24815/siele.v10i2.27237.
- [33] Gonzales, W., Hughes, M. T. Leveraging a Spanish Literacy Intervention to Support Outcomes of English Learners. *Reading Psychology*. 2021;42:411-434. doi:10.1080/02702711.2021.1888352.
- [34] Chen, M. The Impact of Mobile Learning on the Effectiveness of English Teaching and Learning-A Meta-Analysis. *IEEE Access*. 2022;10:38324-38334. doi:10.1109/ACCESS.2022.3165017.
- [35] Tarazi, A., Arafat, S. The Role of Using iPad Technology in Enhancing Students' Motivation toward Learning the English Language. *EDU REVIEW. Revista In-ternacional de Educación y Aprendizaje*. 2021. doi:10.37467/gka-revedu.v9.2971.
- [36] Garzón, J., Lampropoulos, G., Burgos, D. Effects of Mobile Learning in English Language Learning: A Meta-Analysis and Research Synthesis. *Electronics*. 2023;12(7). doi:10.3390/electronics12071595.
- [37] Saito, K., Petrova, K., Suzukida, Y., Kachlicka, M., Tierney, A. Training auditory processing promotes second language speech acquisition. *Journal of experimental psychology. Human perception and performance*. 2022; 48(12): 1410- 1426. doi:10.1037/xhp0001042.
- [38] Shi, Y., Tsai, C. A Canonical Correlation Analysis: Learning Preferences and Sensory-Specific English Learning Material Stimuli. *Revija za elementarno izobraževanje*. 2022. doi:10.18690/rei.15.4.493-510.2022.
- [39] Ariani, S., Valiantien, N., Rachmawaty, N. STUDENTS' LANGUAGE LEARNING STYLE PREFERENCES AT ENGLISH LITERATURE STUDY PROGRAM. *CaLLs (Journal of Culture, Arts, Literature, and Linguistics)*. 2021. doi:10.30872/CALLS.V7I1.5160.
- [40] Liu, J., Liu, X., Yang, C. A study of college students' perceptions of utilizing automatic speech recognition technology to assist English oral proficiency. *Frontiers in Psychology*. 2022;13. doi:10.3389/fpsyg.2022.1049139.
- [41] Wu, X., Zhang, Y., Zhu, W. Study on an English Speaking Practice System based on Automatic Speech Recognition Technology. *Journal of Education and Educational Research*. 2023. doi:10.54097/jeer.v4i1.10273.
- [42] Wu, X., Zhang, Y., Zhu, W. Study on an English Speaking Practice System based on Automatic Speech Recognition Technology. *Journal of Education and Educational Research*. 2023. doi:10.54097/jeer.v4i1.10273.
- [43] Liu, J., Liu, X., Yang, C. A study of college students' perceptions of utilizing automatic speech recognition technology to assist English oral proficiency. *Frontiers in Psychology*. 2022;13. doi:10.3389/fpsyg.2022.1049139.

- [44] Ajmal, M. Exploring the Role of Motivation in English Language Teaching: Learners and Teachers Perspective. 2021. doi:10.17762/PAE.V58I1.804.
- [45] Selimovic, S. An Exploration of Students' Motivation in Learning English. 2022. doi:10.53880/2744-2454.2022.2.2.24.
- [46] Santovac, D. M. P., Popović, A. V. H. INTRINSIC AND EXTRINSIC MOTIVATION IN EFL LEARNING AT UNIVERSITY LEVEL. 2022. doi:10.21618/fil2225086p.
- [47] Mahmud, S., Akmal, S., Arias, A. Is it More Intrinsic or Extrinsic? The Motivation of Gayonese EFL Students to Learn English. 2023. doi:10.26811/peuradeun.v11i1.816.
- [48] Jiang, M., Jong, M., Lau, W., Chai, C., Wu, N. Using automatic speech recognition technology to enhance EFL learners' oral language complexity in a flipped classroom. Australasian Journal of Educational Technology. 2021;37:110-131. doi:10.14742/AJET.6798.
- [49] Kochem, T., Beck, J., Goodale, E. Use of ASR-Equipped Software in the Teaching of Suprasegmental Features of Pronunciation. CALICO Journal. 2022. doi:10.1558/cj.19033.
- [50] Hanh, N. H., & Huyen, P. M. (2024). A Review of Collaborative Learning Activities to Help Students Reduce Speaking Anxiety in EFL Classroom Settings. American Research Journal of Humanities & Social Science (ARJHSS), 7(03), 86-90.
- [51] Suess, N., Hauswald, A., Zehentner, V., Depireux, J., Herzog, G., Rösch, S., Weisz, N. Influence of linguistic properties and hearing impairment on visual speech perception skills in the German language. Plos one, 17(9), 2022:e0275585.
- [52] Gray, A. T., Macpherson, L., Carlin, F., Sossen, B., Richards, A. S., Kik, S. V., ... Esmail, H. Treatment for radiographically active, sputum culture-negative pulmonary tuberculosis: A systematic review and meta-analysis. Plos one, 18(11), 2023:e0293535.
- [53] Mirheidari, B., Bittar, A., Cummins, N., Downs, J., Fisher, H. L., Christensen, H. Automatic detection of expressed emotion from Five-Minute Speech Samples: Challenges and opportunities. Plos one, 19(3), 2024:e0300518.

Attachment I Questionnaire

Category	Item
Basic Information	1. Sex
	2. Age
	3. Education
	4. English Proficiency
	5. Frequency of daily use
	6. Device level
Metropolitan Background	1. What type of city do you currently reside in?
	2. What type of Apple device do you currently use for English learning?
	3. How do you think living in a metropolitan city affects your English learning?
Travel Abroad Experience	1. Approximately how many times do you travel abroad in a year?
	2. How often do you use English during your travels abroad?
	3. Do you think traveling abroad has improved your English-speaking ability?
Device Type and Usage Frequency	1. How frequently do you have opportunities to engage with and use English?
	2. What type of Android device do you currently use for English learning?
	3. How would you rate the performance of your device when using ASR technology for English learning?
Device Satisfaction	1. How satisfied are you with your device for English learning overall?
	2. Does your device's performance meet your learning needs when running ASR applications?
	3. Approximately how many hours per week do you use your device for English learning?
Visual Learning Preference	1. When learning new English vocabulary, I tend to:
	2. When understanding English grammar, I find it easier to learn more through:
	3. If I need to memorize an English speech or dialogue, I prefer:
Auditory Learning Preference	1. I prefer learning English listening skills through:
	2. When trying to understand complex English expressions or idioms, I tend to:
	3. When learning English pronunciation, I prefer:
ASR Technology Usage	1. How frequently do you use ASR technology to practice English speaking?
	2. How user-friendly do you find ASR technology for English learning?
	3. Are you satisfied with the effectiveness of ASR technology in improving your English-speaking ability?
	4. How well does ASR technology perform in recognizing the accuracy of your pronunciation?
	5. How much impact do you think ASR technology has on your English learning experience?
Learning Efficiency	1. On average, how much time do you spend on English learning each day?
	2. Compared to other learning tasks, how fast do you complete English learn
	3. How would you rate your English learning efficiency?
	4. When using ASR technology for English learning, how do you feel about your learning efficiency?
	5. How quickly do you achieve your learning goals during the English learning process?
Learning Motivation	1. I learn English mainly because I enjoy the process of learning a language.
	2. I believe learning English is important for my personal development.

	3. When I learn English, I feel a sense of accomplishment.
	4. Mastering English makes me feel more confident.
	5. I learn English because it can improve my career prospects.
	6. If it weren't for work or school requirements, I wouldn't learn English.
	7. I learn English to better communicate with people from different countries.
	8. I learn English mainly to pass exams or obtain certifications.
Improved English Speaking Ability	1. What score does ASR technology typically give when assessing your English speaking?
	2. After using ASR technology, do you feel your English-speaking accuracy has improved?
	3. How would you rate your English-speaking ability?
	4. Compared to before using ASR technology, how has your speaking fluency changed?
	5. Based on your observations, how do others evaluate your English-speaking ability?