RESEARCH
www.commprac.com
ISSN 1462 2815

AUTOMATIC SPEECH RECOGNITION INDAH (ASRI) FOR ENHANCING STUDENTS' PRONUNCIATION

Indah Muzdalifah [1], M. Zaim [2]* and Refnaldi [3]

[1] The Doctoral Student of Universitas Negeri Padang, Padang, Indonesia.
[2,3] Universitas Negeri Padang, Padang, Indonesia.
*Corresponding Author Email: mzaim@fbs.unp.ac.id

DOI: 10.5281/zenodo.11090709

Abstract

This research aims to change the conventional learning paradigm into a pleasant, practical learning method**.** This research uses ADDIE Research and Development. Instruments use interview guides, questionnaires, tests, and observations. This study involved 58 (control classes and experiments. The results of this research are, first, students' need, lack, and want for learning media based on the ASRI website. Second, the results of product validity on the ASRI website were declared valid it is 78.6% of the results of product practicality were concluded to be very practical or 90%, and the effectiveness pre-test (54.2) and post-test (88.6) were very practical. Third, the hypothesis results show a significant difference in the usage of ASRI media in pronunciation learning. The contribution of this research is (1) the ASRI website is interactive; (2) it provides feedback (notification); (3) it is especially (designed); (4) it is accurate; (5) and flexible (handy) or INDAH.

Keywords: Automatic Speech Recognition, Pronunciation, Learning Media, Research and Development.

INTRODUCTION

Pronunciation is one of the significant components of language learning. The cornerstone of communication is clear and understandable speech (It, 2023). The goal of verbal pronunciation assessment is to create a tool for students education overseas languages like English to correct their pronunciation automatically (Shufang, 2021). Students may comprehend the features of English articulation from hearing distance, image, and drop and enhance their English articulation (Zhou & Zhao, 2022). As a result, it is critical to develop a systematic and practical automated correction solution for English pronunciation issues (Sheng & Yang, 2021). The intermediary sheet between audio and linguistic representations is pronunciation lexicons. Nowadays position in speech recognition, the quality of the pronunciation component determines the total system presentation (Yolchuyeva et al., 2019)

By using technology, students can more easily learn pronunciation. Concentrating on computer-aided pronunciation training is beneficial for language teaching since it involves language acquisition's fundamental cognitive mechanisms (Hamilton et al., 2024). Since it is the most intuitive way of engagement for humans, the primary purpose of speech recognition is to allow persons to cooperate more efficiently and proficiently (Jambi et al., 2022). There is controversy concerning the particular definition of digital learning utensils (specific informative technologies) and e-learning atmospheres (more extensive backgrounds combine technological, learning, and pedagogical components) (Lin & Yu, 2023). One of the critical essential knowledges of trade information is linking it to the numerical world via wireless communication devices (Yang & Lv, 2021). Since technology has advanced fast in recent years, having a substantial impact on many aspects of human lives, as well as education, using expertise in the informative procedure can provide fresh views and opportunities for learning (Polyzi & Moussiades, 2023)

Computer Aided Language Learning (CALL) is commonly used as a computer-aided learning medium**.** CALL systems are predicted to participate significantly in language regeneration (Ito, 2023). CALL is finding and learning computer applications in language teaching and learning (Shamshiri et al., 2023). Besides, CALL supports the progress in the English ability of persons in dissimilar areas (Korzekwa et al., 2022). Today, with computers playing an increasingly crucial role in society activities, computer competitions container play an essential part in education and strengthen previous interpretations (Polyzi & Moussiades, 2023). It is thought that using intelligent computer systems can help scientific students overcome their language issues. On the other side, this can impact their learning process while providing a solution to language teachers' challenges (Chaib et al., 2023). Furthermore, while several large language CALL systems include speaking capabilities such as pronunciation and discussion exercise, it is uncommon for a threatened CALL arrangement to provide such functionality (Ito, 2023).

One form of computer-assisted language learning that can be utilized in pronunciation learning is computer-assisted pronunciation technology (CAPT). CAPT has improved language learners' learning results, leading to increasing research on this topic (Cengiz, 2023). In EFL instruction, CAPT has been utilized to assist students improve their articulation (Sun, 2023)**.** CALL and CAPT systems may be customized, and one way to make them more interesting is to put them on games or robots, particularly for older children (Farrús, 2023). Teachers play a crucial role in promoting the effectiveness of CAPT by scaffolding, motivating, and providing feedback to learners (Cengiz, 2023). Computer-assisted pronunciation training has been used to target specific aspects of pronunciation, such as vowel perception, with promising results (Thomson, 2011). (ASR) is one of the greatest widely utilized CAPT technologies, capable of decoding and transcribing people's speech and providing pronunciation feedback (Cengiz, 2023)**.**

One of the CAPT applications that can be utilized in Pronunciation knowledge is ASR. On the other hand, ASR is a form of CAPT organization, a speech-decoding system and transcribing technology that allows students to learn subject independently (Sun, 2023). Thus, ASR is a technique that empowers computers to transform human voice into writing. ASR was created utilizing a variety of methodologies, numerical representations, engine knowledge, and neural systems are just a few examples (Sun, 2023). ASR is rapidly employed in emergency response centres, home voice assistants, and search engines. Because spoken language is so important in our lives, it is crucial that ASR systems can deal with the variety in how people talk (Feng et al., 2024). On the other hand, other names for it include Speech-to-type (STT), (ASR), and computer voice recognition (CVR). It consists of computer engineering, languages, technical expertise, and research (Jambi et al., 2022)

In fact, in the field, students often make pronunciation mistakes that cause a shift in meaning**.** People's communication is built on clear and intelligible speech. The expression of meaning is directly affected by whether or not the pronunciation is correct (It, 2023). Feedback has been suggested to be one of the essential learning methods across all kinds of instruction, whether classroom teaching or remote learning, and for developing all language skills, such as reading, writing, and speaking (Zhang & MacWhinney, 2023). As technology has advanced fast in recent years, impacting many aspects of society education, incorporating the incorporation of expertise into the informative progression can provide fresh views and opportunities

for education (Zhang & MacWhinney, 2023). CAPT technologies may detect and diagnose pronunciation faults in learners and assist them in correcting these issues (Cengiz, 2023). As a result, research into programmed articulation mistake discovery has applied implications. The most common sorts of spoken pronunciation faults include phoneme error, lost pronunciation, pronouncing more pronouncing, and pronouncing time errors (Zhou & Zhao, 2022). For the English pronunciation mistake correction system, the degree of corrective presentation and the dependability, attainability, and flexibility of evidence response are the primary criteria for assessing its brilliant overall presentation (Sheng & Yang, 2021). Feedback from iCALL-based speaking activities may result in more correct utterances (Jiang et al., 2022).

The second problem that often occurs in the field is that lecturers give examples of their pronunciation, causing inaccuracies in pronunciation. For the English pronunciation mistake correction system, the degree of corrective presentation and the dependability, workability, and adaptability of data reaction are the primary criteria for assessing its outstanding overall presentation (Sheng & Yang, 2021). As an aid, incorporating technology initiates and successfully improves the proper comprehension of phonemes by artificial and replication of the misprinted dictionary instantly (Sattar & Alghamdi, 2023)

The third problem was that lecturers taught pronunciation with makeshift media such as dictionaries containing symbols. The technique of converting voice into a series of words using computer software is recognized as automatic speech recognition (Zhang & MacWhinney, 2023). Automatic speech recognition (ASR) can benefit dysarthric speakers (Qian et al., 2023). The application of ASR knowledge in students' knowledge has been shown to aid students by improving energetic subsequent language oral practice, such as oral grammar and pronunciation (Bashori et al., 2021).

The next problem often occurs when students are given only 2x50 minutes in 1 week in pronunciation learning on campus, so they experience a lack of time in practice. Mobile technology usage in education has been a trend in recent years. The younger generation nowadays is born and grows up in a society where digital devices are used in both the home and the kindergarten setting (Pei et al., 2023). Computers are unquestionably the most critical and complicated technical gadgets, and computer-aided learning has become a primary emphasis for unconventional self-learning settings (Hanci-Azizoglu & Vural, 2023)**.** MALL refers to the study of a second or a foreign language with the use of cell phones, tablets, and personal digital assistants are examples of mobile devices (Liu et al., 2022). ASR is a potential enlightening technology in MALL that can give learners a straightforward approach to learning languages and practicing pronunciation (Liu et al., 2022)

The next problem is that pronunciation learning is considered boring because students only listen to explanations from lecturers and imitate pronunciation from lecturers. "Web-based Tools and Methods for Rapid Pronunciation Dictionary" examines using web-derived pronunciations to create or improve current pronunciation dictionaries (Tim Schlippe, Sebastian Ochs, 2014). Applying computers as an instrument in education makes it unique in student interaction. Through this interaction, learners can develop their language proficiency (Shamshiri et al., 2023). Because it is the most intuitive way of engagement for people, the primary purpose of speech recognition is to allow them to cooperate more readily and proficiently (Jambi et al., 2022). CALL

has several interactive aspects that help in the appearance, strengthening, and calculation of the subject to be educated (Shamshiri et al., 2023)

This research aims to change the conventional learning paradigm system, generally dominated by lecturers (teacher centre), into a more enjoyable practical learning method because students and lecturers can interact on the ASRI website. This research is essential because ASRI's website-based media can correct pronunciation errors because there are notifications as student pronunciation feedback so that students know where the sound errors are produced. In addition, the ASRI website can improve proper accuracy through the examples contained in it.

The formulation of the problem in this study is:

1. What kind of learning media do students need?
2. What is ASRI's website media products' validity, practicality, and effectiveness?
3. What is the research hypothesis for using the ASRI website's pronunciation learning media?

METHOD

This investigation is a study and expansion research. This study involved 58 students (1 control class and 1 experimental class) of FKIP English Universitas Pahlawan. This research is a research development research that practices the ADDIE model (analysis, design, development, implementation, evaluation). This R&D project seeks to create a media-based website that is ASR in learning pronunciation. The development model was built on the ADDIE paradigm, which has five stages: need analysis, design, development, implementation, and evaluation

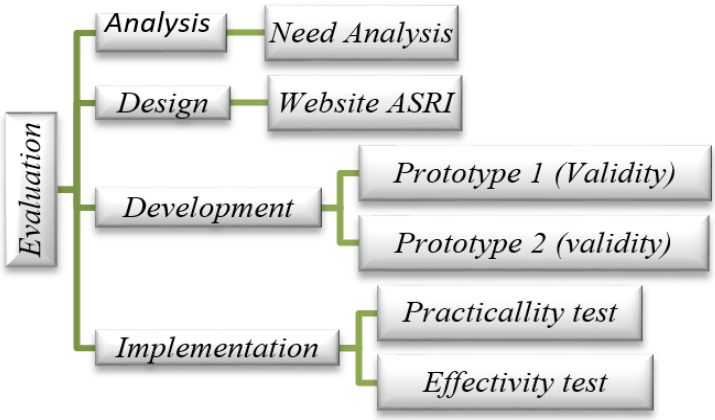

Figure 1: Stages of ADDIE Model

Questionnaires, observations, sheets, and tests were used as study instruments. Data were analyzed descriptively, followed by an examination of the validity, practicability, and effectiveness of the ASRI media. The subjects of this study are divided into two groups: control (29 students) and experimental (29 students). The need analysis data obtained were analysed in the Likert Scale assessment category. The product's effectiveness (ASRI website) was tested using normality, homogeneity, and hypothesis tests. Data were descriptively evaluated to describe the validity and practicality of the developed media. SPSS 26 was used to examine the effectiveness data in the form of student learning outcomes.

RESULTS AND DISCUSSION

Need Analysis of Learning media

The first finding in this study is to answer the research formulation "what type of learning media are needed by students?". To find this answer, data collection was carried out in the form of questionnaires which were distributed through online questionnaires to 29 students. So that the following data are obtained. Based on need analysis data (results of interviews with lecturers and student questionnaires), it was found that students need learning media based on the ASRI website; students still lack learning media based on the ASRI website; students want learning media based on the ASRI website.

Table 1: Students Need Analysis

Indicator		Need			Lack			Want		
		Total Score	Mean Score (%)	Result	Total Score	Mean Score (%)	Result	Total Score	Mean Score (%)	Result
I	Interactive	139	84,8	Desperately need	128	78	Lack	137	83,5	Eager to
		139	84,8	Desperately need	127	77,4	Lack	140	85,4	Eager to
		140	85,4	Desperately need	129	78,7	Lack	138	84,1	Eager to
		138	84,1	Desperately need	128	78	Lack	139	84,8	Eager to
N	Notification	142	86,6	Desperately need	130	79,3	Lack	141	86	Eager to
		140	85,4	Desperately need	132	81	Very lacking	135	82,3	Eager to
		133	81,1	Desperately need	130	79,3	Lack	141	86	Eager to
		137	83,5	Desperately need	133	81	Very lacking	137	83,5	Eager to
D	Design	137	83,5	Desperately need	131	80	Very lacking	138	84,1	Eager to
		140	85,4	Desperately need	133	81	Very lacking	138	84,1	Eager to
		139	84,8	Desperately need	130	79,3	Lack	139	84,8	Eager to
		138	84,1	Desperately need	131	80	Very lacking	138	84,1	Eager to
A	Accurate	142	86,6	Desperately need	133	81	Very lacking	141	86	Eager to
		143	87,2	Desperately need	132	81	Very lacking	140	85,4	Eager to
		141	86	Desperately need	131	80	Very lacking	138	84,1	Eager to
		143	87,2	Desperately need	134	81,7	Very lacking	143	87,2	Eager to
H	Handy	142	86,6	Desperately need	131	80	Very lacking	142	86,6	Eager to
		138	84,1	Desperately need	131	80	Very lacking	142	86,6	Eager to
		139	84,8	Desperately need	129	78,7	Lack	139	84,8	Eager to
		140	85,4	Desperately need	131	80	Very lacking	141	86	Eager to

Based on the study's findings, the media required in learning pronunciation is media in the form of a website with the goal of making it easier for students to interact with the learning that has been arranged on the website so that students can listen and pronounce the correct sound. Second, the media required for learning pronunciation

is a website-based media with the ability to translate sounds or notifications that provide information on right pronunciation. Third, website-based media that are specially built in a systematic and sequential manner based on needs in order to satisfy the student's characteristics design are essential in learning pronunciation. An excellent medium should be clear and inviting.

Design of Website ASRI

Innovative education media design is a learning methodology that employs tools or programs that are not traditional in nature. If traditional learning methods are applied, students will be less interested and motivated to participate in learning activities. Students as a result have poorer learning results and fewer engaging learning activities and experiences. Learning should ideally be contextual and authentic, because classroom knowledge is frequently artificial and detached from students' daily concerns in real life. Student-centered learning is emphasized in innovative learning with teaching media. The ASRI website media is the product design in this development investigation. This research designed the ASRI website media based on five characteristics that were obtained from needs analysis, it is INDAH (Interactive, Notification, Design, Accurate, and Handy).

The ASRI website media is built on interactive media characteristics. A learning media must have clear learning objectives and be simple to use. In this case, the ASRI website's learning media design attempts to make it simple for students to interact with the learning that has been structured on the ASRI website so that students can accurately listen to and pronounce words. The ASRI website was designed with quality and precision in mind. In this case, the ASRI website's learning media design has the quality of a system capable of translating sounds or notifications to provide information on right pronunciation. The media on the ASRI website is designed specifically for students. The ASRI media website is composed of five syntactic elements.

Development Website ASRI

This development research is divided into five learning activities: identification, notification, decision, application, and harmonization INDAH. First syntax is Identifying. It identifies each segmental (consonant, vowel, and diphthong) and suprasegmental (stress and intonation) pronunciation learning resource. The second syntax is Notification, which is responsible for notifying users about segmental (consonant, vowel, and diphthong) and suprasegmental (stress and intonation) pronunciation learning information. The third syntax is Decision, a learning action in which students are asked to determine each phoneme of segmental (consonant, vowel, and diphthong) and suprasegmental (stress and intonation) pronouncing learning material. The fourth syntax is Application, in which students can customize the ASRI app to practice segmental (consonant, vowel, and diphthong) and suprasegmental (stress and intonation) pronunciation. Harmonization is the final syntax. Students' ability in segmental (consonant, vowel, and diphthong) and suprasegmental (stress and intonation) pronunciation is examined in this part

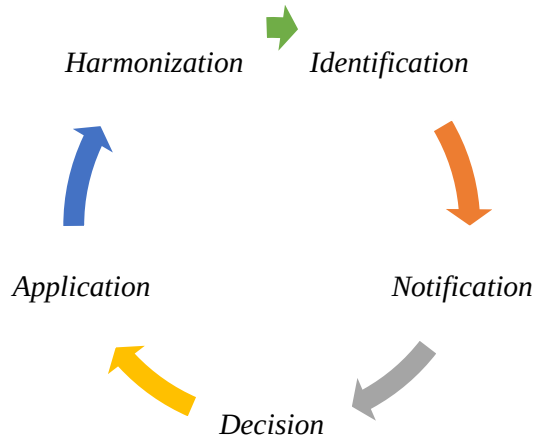

Figure 2: ASRI Media Learning Syntax

Based on the previously outlined theoretical and practical studies of media design, the design must be produced in accordance with the validity assessment from expert opinion. A prototype is used in development research as a product development process before being used to learning

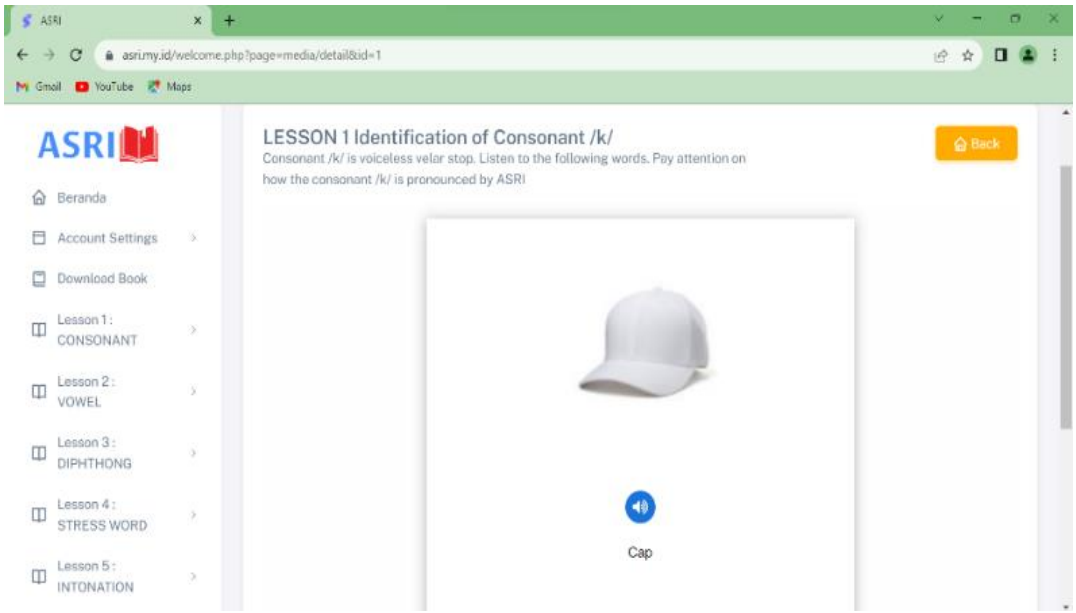

Figure 3: Identification Step

The image is the initial section of the identifying screen. This identification area has numerous features, including graphics that illustrate a word whose sound you want to know. The image above depicts the sound "cap" along with a picture of a hat. Below the image is a speaker button that displays a symbol where the user can hear the sound of the word hat. There is a "turn" button at the bottom of the speaker button, which allows pupils to return to the main menu. There is a description of card of 24 under the "turn" function, indicating that the word "hat" is the first of the 24 words that pupils will learn

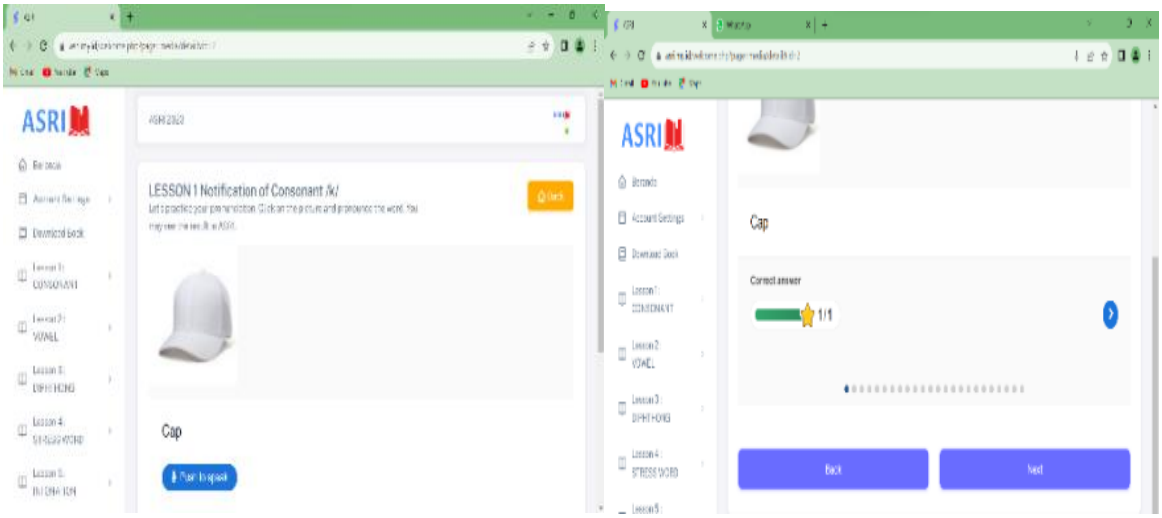

Figure 4: Notification Step

The image depicts the second stage, Notification, in which pupils are given the opportunity to pronounce words in response to the images and phrases displayed on the screen. The ASRI website can determine the pronunciation of this word by pressing the "push to speak" button. Students can then repeat their pronunciation

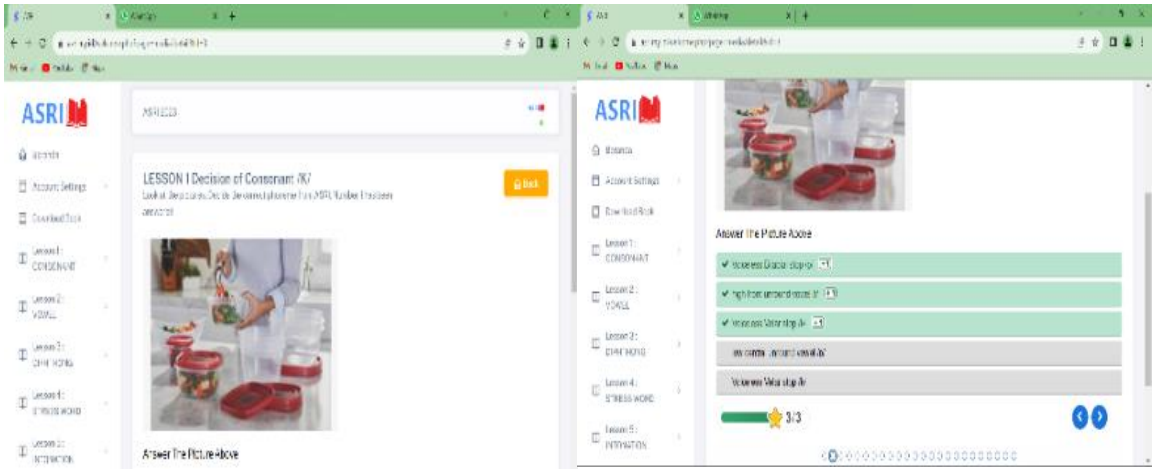

Figure 5: Decision Step

Those images are displayed on the screen above to help pupils remember the preceding words in the identification and notification portions. This is due to the fact that all images or materials used between identification, notification, and decision are the same. For each image on the screen above, various phoneme answer options are displayed. Students are instructed to choose the correct phoneme based on the words offered. When students decide, the ASRI website responds to the answer as displayed on the screen. Then, at the bottom, there are 24 dots, indicating that pupils must complete 24 questions. These 24 questions are broken into three sections: the beginning sound, the middle sound, and the final sound. Then, on the right, there are "next" buttons

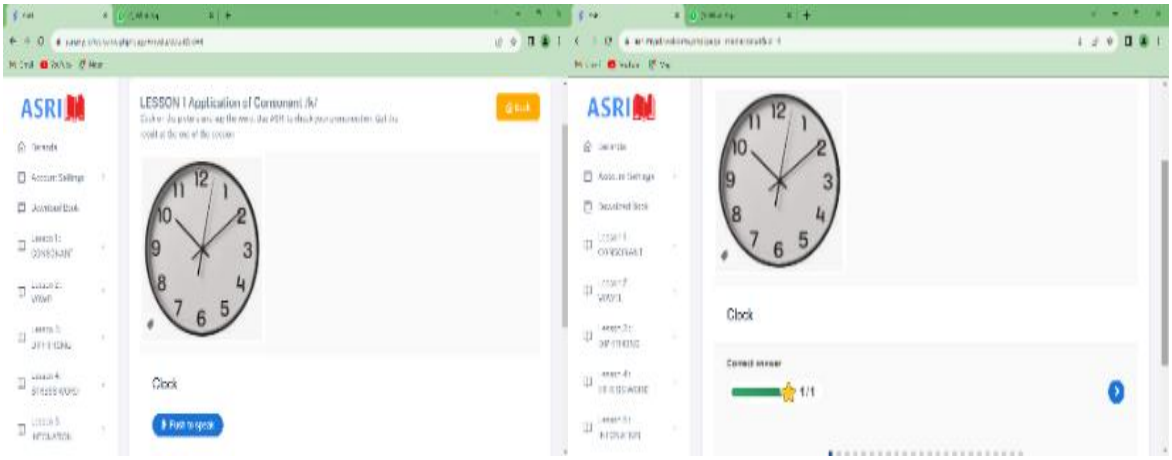

Figure 6: Application Step

Based on the image, it is mentioned in the Application section that pupils are presented with an image, followed by words that describe the visual. The button "push to speak" is displayed beneath the phrases, implying that pupils are instructed to use the button in order to apply the words provided. Students are instructed to wait for a response from the ASRI website. The website processes the detected sound into the program, as seen on the screen shortly after the button is pushed. Students must wait a few moments for the outcomes of the noises they have entered. The website has replied to the student's voice input, as seen on the screen. If it is correct, it will appear on the "correct" list

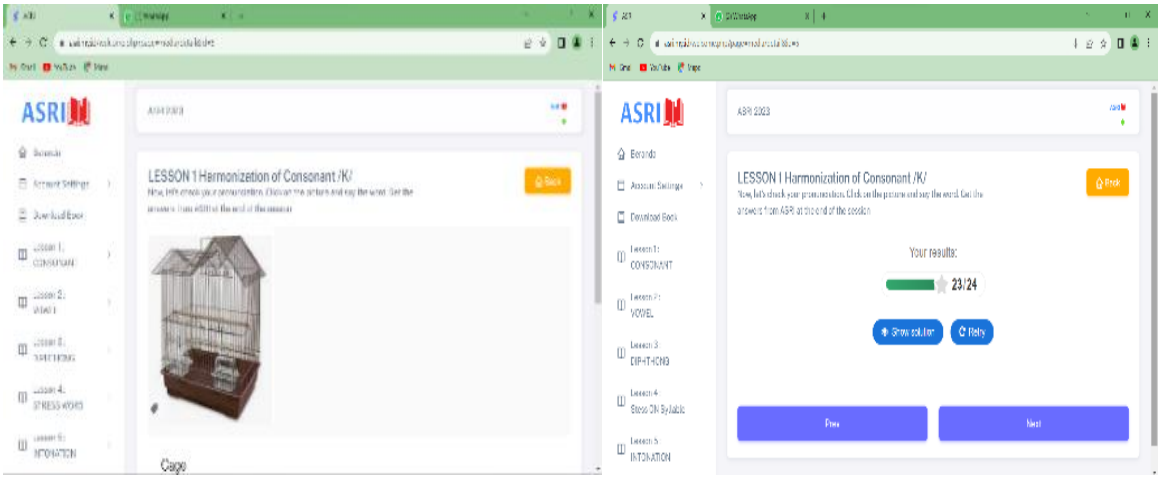

Figure 7: Harmonization Step

Those images are displayed on the screen for students to analyze. The image also includes a word beneath it, so pupils are prepared to recite the word on the button beneath it. The screen displays a "push to speak" button (similar to the one in the Application section). Students can begin pronouncing the appropriate words so that the website can recognize and process their voices. There is also a 24 dot symbol at the bottom. These 24 questions are broken into three sections: the beginning sound, the middle sound, and the final sound. On the screen, you can see the website detecting and processing the student's sound input. Students are instructed to wait for the screen to respond in the manner depicted in the following image.

Implementation Website ASRI

The use of ASRI media research products on research objects has been evaluated. This series of studies served as an initial step in determining the experimental class, which received therapy in the form of learning books and ASRI media, and the control class, which did not get treatment. The outcomes of this implementation will affect the outcomes of future research studies, specifically the product's effectiveness and practicality. The fourth step in development research is implementation. The application or implementation of research is known as implementation. Implementation entails giving a mechanism to carry out something, as well as having an effect or result on anything. The implementation phase marks the end of the development activities and the formative evaluation system with research into the creation of ASRI materials and media, which has determined that at this implementation stage, the product is capable of having an impact on pronunciation learning as a result of the usage of ASRI-assisted materials and media.

The first step in the implementation stage of this development research is to identify the research object by identifying the class that is given treatment, which is then referred to as the experimental class, and the class that is left alone in pronunciation learning, which is referred to as the control class. The previous description stated that the sample data was declared homogeneous and had the same background between the control and experimental classes, so it was appropriate to continue making comparisons between the control and experimental classes.

The sample size for these two classes was the same, 29 students. The lecturer will then administer a pre-test to both classes at the first meeting of learning activities so that they may compare before and after the application of ASRI materials and media. The following step is to observe student and lecturer actions in the classroom, in both control and experimental classes. The speaker uses ASRI materials and media in the experimental class. At the second meeting, the lecturer introduced students to the ASRI website by offering instructions and activities to be completed during the learning process.

Evaluation website ASRI

The outcome of the pre-test and post-test for the experimental class and the control class pre-test and post-test are presented in the table above. The experimental group scored 54.2 on the pre-test and 88.6 on the post-test, while the control group scored 54.2 on the pre-test and 55.3 on the post-test after the pre-test and post-test scores for the control and experimental groups. As a result, the following tests are required: (1) statistical analysis data, (2) normality test, (4) homogeneity test, and (5) hypothesis test.

Validity: the product validity result in the form of the ASRI website is declared valid, which is 78.6%. Practicality: the results of the product practicality questionnaire in the form of the ASRI website, which students assessed, were concluded to be in the convenient category of 90%.

Effectiveness: Based on data from quasi-experimental design trials, it was found that students' scores in experimental classes increased significantly amid the pre-test (54.2) and post-test (88.6). Based on the hypothesis test data, it is proven that there is an influence on the use of pronunciation learning media using ASRI media compared to articulation defence that does not use ASRI media.

Statistical analysis data

This training employed a quasi-experimental design research approach. To gather data in groups (experimental and control), tests/questions (pre-test and post-test) are used. The experimental group is exposed to ASRI media, whereas traditional media is shown to the control group. The next are the results of descriptive analysis data testing.

Table 2: Data Analysis

Pre-test Experiment	N	Minimum	Maximum	Mean	Standard Deviation
Pre-test Experiment	29	39	69	54.24	8.153
Post-test Experiment	29	77	98	88.66	6.548
Pre-test control class	29	39	69	54.24	8.153
Post-test control class	29	40	69	55.31	8.181
Valid N (listwise)	29				

The sample size is 29 based on the information presented above. The experimental class pre-test scores range from 39 to 69, with a mean of 54.24 and an SD of 8.15. The experimental post-test group had scores ranging from 77 to 69, with an average of 88.66 and a standard deviation of 6.548. Third, the score range for the control class pre-test was 39 to 69, with an average of 54.24 and an average nonconformity of 8.15. Fourth, the control class post-test received a minimum score of 40 and a maximum score of 69, with a normal score of 55.31 and a normal deviation of 8.18.

Normality Test

The study data is then submitted to a normality test to determine whether or not it is usually spread. Standard data is essential previously doing parametric statistics (in this example, the paired sample t test and independent-sample t-test). Two normality tests are commonly used in parametric statistics: the Kolmogorov-Smirnov test and the Shapiro-Wilk test. The normality data shown below was gathered.

Table 3: Tests of Normality

Class	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Pre-test experiment (media ASRI)	.139	29	.162	.955	29	.241
Post-test Experiment (media ASRI)	.133	29	.200*	.920	29	.030
Pre-test control (conventional)	.139	29	.162	.955	29	.241
Post-test control (conventional)	.142	29	.138	.931	29	.058

*. This is the actual significance's lower bound.
a. Lilliefors Significance Adjustment

Based on the results, the consequence value (sig) for all data in the Kolmogorov - Smirnov and the Shapiro-Wilk tests is more than 0.05, indicating that the research data is regularly distributed. Parametric statistics (paired t-test and independent sample t-test) can be used to assess the study data since it is regularly distributed.

Homogeneity Test

The homogeneity test controls if the data adjustment from two or more groups is homogenous (the same) or heterogeneous (different). Homogeneous data is one of the prerequisites for the independent sample t-test (but it is not required). The homogeneity test is utilized in this study to examine whether or not the variance of the experimental class post-test data (ASRI media) and the control class (conventional) post-test data is homogenous.

Table 4: Test of Homogeneity of Variance

Students Result in Study	Levene Statistic	df1	df2	Sig.
Based on Mean	.590	1	56	.446
Based on Median	.634	1	56	.429
Based on the Median and with adjusted df	.634	1	52.638	.429
Based on trimmed Mean	.647	1	56	.424

Structure the homogeneity test was performed in this study with a significance of > 0.05 using SPSS 21 software, and the following was the foundation for decision-making:

- If the significance or probability value is 0.05, the variance of two or more data groups is considered heterogeneous.
- If the significance or probability value is more significant than 0.05, the variance of two or more data groups is said to be homogenous.

The significant value (sig) Based on the Mean is based on the above result 0.446 > 0.05, implying that the experimental and control class post-test data variance is the same or homogeneous. As a result, one of the independent sample t-test criteria has been met.

Paired Sample t-Test

The following step is a paired sample t-test to see if the average of two paired samples varies. The paired sample t-test assumes that the data is regularly distributed. Like the prior data, the distribution data is stated to be expected. This study employed the matched sample test to see whether there was a difference in an average increase in pronunciation skills after hearing lectures utilizing ASRI learning material. The paired sample t test was used to the experimental class pre-test and post-test data (using ASRI media) to test this hypothesis.

Table 5: Paired Samples Test

Paired Differences							t	df	Sig. (2-tailed)
Mean			Std. Dev.	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Pre-test Exp- Post-test Exp	-34.414	10.805	2.006	-38.524	-30.304	-17.151	28	.000
Pair 2	Pre-test Control - Post-test Control	-1.069	10.957	2.035	-5.237	3.099	-.525	28	. 000

Based on the output of Pair 1, a Sig (2-tailed) value of 0.000<0,05 is produced, implying that there is a difference in average student learning outcomes between the experimental class pre-test and the experimental class post-test (ASRI media). The Sig value is calculated using the output of pair 2. (0,000 <0,05, 2 tails). As a result, it is possible to conclude that there is a difference in average student learning outcomes between the control class pre-test and the control class post-test.

Hypothesis Test

Apply the t-test to assess whether the regular participant learning results in the experimental class using Automatic Speech Recognition materials and media are meaningfully sophisticated than in the control class. The hypothesis proposed is "Is there any significant effect of using ASRI-assisted materials and media on EFL students' pronunciation learning outcome?" The paired sampler t test was done on the experimental class pre-test data with the experimental class post-test (ASRI materials and media) to answer the query. A Sig (2-tailed) value of 0,000<0,05 is produced from the output of Pair 1. As a result, the pre-test and post-test average student learning results for the experimental class differ (ASRI media). When the formula mentioned above is applied, H1 is authorized, whereas H0 is rejected. In other words, the characteristics/measures of average student learning outcomes vary. So, based on the discussion of output pair 1, the conclusion that can be reached is that it can solve the issue from the problem formulation, and ASRI learning materials and media have an influence.

DISCUSSION

Based on the results of the research obtained, it was found that there was a sharp difference between pronunciation learning using ASRI website media and conventional learning that did not utilize technology as a pronunciation learning medium. The current generation has been spoiled with the convenience provided by technology in every aspect of life. This is what makes the background of researchers create a tool that can be used in one of the learning, it is pronunciation. This has been proven by several previous researchers who found the use of website-based pronunciation learning media can help students in pronunciation learning. Student pronunciation in English can be more precise and better by utilizing ASRI website-based media to help students find (reconcile) the pronunciation mistakes made (Cengiz, 2023; Zhou & Zhao, 2022). The advantage of using a website in pronunciation learning is that this system is equipped with student needs related to incorrect responses or correct pronunciation. The students need media Automatic Speech Recognition that has characteristic a website-based media with the ability to translate sounds or ***notifications*** are required to offer information on pronunciation correctness. ASRI website has a feature that can respond incorrectly or correctly to student pronunciation. The research product offered is a website that can provide feedback for its users so that it can provide pronunciation corrections (It, 2023; Zhang & MacWhinney, 2023).

The limited space and time experienced by students while on campus can be overcome by using ASRI website-based media learning pronunciation so that learning does not have to be done in the classroom but can be accessed anytime and anywhere (mobile/ handy). ASR can help learners learn phonology and accent in innovative ways; for example, several programs can 'listen' to a learner's pronunciation and provide formative assessment and feedback on pronunciation correctness. *Last,* the students need media Automatic Speech Recognition that has characteristic ***handy***. The simplicity of engagement is one of the benefits of a website. This adheres to the concepts of mobile learning (M-learning). Mobile Assisted Language Learning (MALL) refers to language learning that can be done anywhere and at any time by using a mobile phone. Currently, technology has made it easier for its users, including for the field of Education (Pei et al., 2023). This ASRI website has the advantage of being

able to be accessed via smartphones to break the constraints of space and time (Hanci-Azizoglu & Vural, 2023; Liu et al., 2022)**.**

Furthermore, the students need media Automatic Speech Recognition that has characteristic a website-based media that are specially designed in a structured and sequential manner based on needs in order to meet the student's characteristics ***design*** are required visualization of the media itself in terms of picture, text, colour, typeface, arrangement, and so on, and usability of the media. Then, on the audio side, it can be evaluated using music, dubbing, language, time, and so on. Product specifications explicitly designed for pronunciation learning materials are poured into the ASRI website to assist lecturers in teaching learning materials created on the website (Zhang & MacWhinney, 2023).

The other aspect of the students' needs media Automatic Speech Recognition that has characteristic high level of ***accuracy*** or precision in order to provide feedback to the learner when utilizing the ASRI website. One of the problems found in the background is that the examples given by lecturers are spoken by themselves but have not met the accuracy standards in English. ASR, from the standpoint of a linguistics pedagogical method, can train EFL students' oral abilities and provide possibilities for evaluating pronunciation quality Websites are certainly a high-potential resource for second language learning (Willoughby, 2024). The other advantage of the ASRI website is that the pronunciation material has adjusted to the material students need, which is vocabulary often found in everyday life (Qian et al., 2023). ASR has been found to profit students by attractive energetic second language (L2) talking preparation, such as oral sentence structure and pronunciation (Bashori et al., 2021). Another advantage offered by the ASRI website media is that it can help lecturers provide accurate pronunciation examples because the sound produced by the ASRI website is a native speaker's voice (valid). Besides, the advantage of a website is the ability of the system to allow fellow users to interact to create a pleasant pronunciation learning atmosphere because students can interact on the website (***interactive***) (Jambi et al., 2022; Shamshiri et al., 2023).

Based on the discussion above, if assembled into one word, it can be arranged into notification, handy, design, accurate, and interactive (n, h, d, a, i). When arranged it will form words (I, n, d, a, h). thus, this website was named Automatic Speech Recognition Indah (ASRI). The researchers decide to design learning media based on the Automatic Speech Recognition website, which includes *interactive, notification, accurate, design, accurate, and handy*. It is abbreviated as INDAH, so this website is abbreviated as Automatic Speech Recognition Indah or ASRI.

The ASRI website was designed with quality and precision in mind. In this case, the ASRI website's learning media design includes the capability of a system capable of translating voices or notifications to provide information on right pronunciation. The media on the ASRI website is designed specifically for students. In addition, the ASRI media website is composed of five syntactic phases, abbreviated as INDAH (*Identification, Notification, Decision, Application, and Harmonization*). These five stages of syntax learning were adjusted to a single semester's lesson plan and went through a curriculum analysis approach similar to that used in pronunciation learning.

CONCLUSION

Five things can be concluded from this study: *First*, there are limitations of technology-based learning media that cause an atmosphere that is not conducive and unpleasant in the classroom because lecturers are more dominant in the school (interactive). *Second*, using ASRI website media can provide feedback for users (notifications) because those who assess the pronunciation are non-human systems. *Third*, the ASRI website is deliberately designed specifically according to the needs and background of students based on needs analysis (design). *Fourth*, the ASRI website can detect student pronunciation inaccuracies so that students can imitate accurate pronunciation. *Fifth*, limited space and time are no longer a barrier for students to learn pronunciation because the ASRI website can be accessed via smartphones to be studied anytime and anywhere.

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) Cengiz, B. C. (2023). Computer-Assisted Pronunciation Teaching: An Analysis of Empirical Research. *Participatory Educational Research*, *10*(3), 72–88. https://doi.org/10.17275/per.23.45.10.3

3) Chaib, S. O., Joti, I., & Khoulji, S. (2023). Evaluation of a Computer-Assisted Language Learning System Based on Adaptive Learning Designed for Self-training in Scientific French Language. *International Journal of Information and Education Technology*, *13*(8), 1284–1296. https://doi.org/10.18178/ijiet.2023.13.8.1930

4) Farrús, M. (2023). Automatic Speech Recognition in L2 Learning: A Review Based on PRISMA Methodology. *Languages*, *8*(4), 1–13. https://doi.org/10.3390/languages8040242

5) Feng, S., Halpern, B. M., Kudina, O., & Scharenborg, O. (2024). Towards inclusive automatic speech recognition. *Computer Speech and Language*, *84*(September 2023), 101567. https://doi.org/10.1016/j.csl.2023.101567

6) Hamilton, A., Morgan, S., Harland, K., & Murphy, B. (2024). Embedding masculinities within a gender conscious relational pedagogy to transform education with boys experiencing compounded educational disadvantage. *Pedagogy, Culture and Society*, *00*(00), 1–21. https://doi.org/10.1080/14681366.2024.2301726

7) Hanci-Azizoglu, E. B., & Vural, E. (2023). Computer-Assisted Autonomous Linguistic Learning: A Novel Approach to Second Language Expertise. *SiSal Journal*, *14*(3), 337–382. https://doi.org/10.37237/140306

8) It, K. (2023). The Reform of Pronunciation Teaching in Colleges and Universities by Praat Software From the Perspective of Deep Learning. *International Journal of Web-Based Learning and Teaching Technologies*, *18*(2), 1–18. https://doi.org/10.4018/IJWLTT.325225

9) Ito, A. (2023). Confidence-based Utterance Selection for a Recognizer-free Spoken Dialogue System. *ACM International Conference Proceeding Series*, 481–484. https://doi.org/10.1145/3587716.3587796

10) Jambi, K., Al-Barhamtoshy, H., Al-Jedaibi, W., Rashwan, M., & Abdou, S. (2022). An Empirical Performance Analysis of the Speak Correct Computerized Interface. *Processes*, *10*(3). https://doi.org/10.3390/pr10030487

11) Jiang, M. Y. C., Jong, M. S. Y., Wu, N., Shen, B., Chai, C. S., Lau, W. W. F., & Huang, B. (2022). Integrating Automatic Speech Recognition Technology Into Vocabulary Learning in a Flipped English Class for Chinese College Students. *Frontiers in Psychology*, *13*(July), 1–16. https://doi.org/10.3389/fpsyg.2022.902429

12) Korzekwa, D., Lorenzo-Trueba, J., Drugman, T., & Kostek, B. (2022). Computer-assisted pronunciation training—Speech synthesis is almost all you need. *Speech Communication*, *142*(June), 22–33. https://doi.org/10.1016/j.specom.2022.06.003

13) Lin, Y., & Yu, Z. (2023). Extending Technology Acceptance Model to higher-education students' use of digital academic reading tools on computers. *International Journal of Educational Technology in Higher Education*, *20*(1). https://doi.org/10.1186/s41239-023-00403-8

14) Liu, J., Liu, X., & Yang, C. (2022). A study of college students' perceptions of utilizing automatic speech recognition technology to assist English oral proficiency. *Frontiers in Psychology*, *13*(December), 1–9. https://doi.org/10.3389/fpsyg.2022.1049139

15) Pei, B. C. W., Mon, C. S., & Alezabi, K. A. A. (2023). The adoption of high contrast graphics and primary colours in developing of educational mobile application (Kiriosity) for preschool children in literacy subject. *AIP Conference Proceedings*, *2756*(1), 1–9. https://doi.org/10.1063/5.0149488

16) Polyzi, P., & Moussiades, L. (2023). An artificial vocabulary learning assistant. *Education and Information Technologies*, *28*(12), 16431–16455. https://doi.org/10.1007/s10639-023-11810-9

17) Qian, Z., Xiao, K., & Yu, C. (2023). A survey of technologies for automatic Dysarthric speech recognition. *Eurasip Journal on Audio, Speech, and Music Processing*, *2023*(1). https://doi.org/10.1186/s13636-023-00318-2

18) Sattar, A., & Alghamdi, S. S. (2023). Investigating the Use of E-Dictionaries as Strategy to Improve Speaking Skill through Practical Activities of Precise Phonemes Realization: Case Study of EFL Undergraduate Students of Haripur University, Abatabad University, & Hazara University Mansehra, . *World Journal of English Language*, *13*(8), 182–193. https://doi.org/10.5430/wjel.v13n8p182

19) Shamshiri, F., Shafiee, S., & Rahimi, F. (2023). Effects of Computer-assisted Language Learning (CALL) and Different Interaction Patterns on Vocabulary Development of EFL Learners. *Journal of Language and Education*, *9*(4), 110–127. https://doi.org/10.17323/jle.2023.12093

20) Sheng, Y., & Yang, K. (2021). Automatic Correction System Design for English Pronunciation Errors Assisted by High-Sensitivity Acoustic Wave Sensors. *Journal of Sensors*, *2021*. https://doi.org/10.1155/2021/2853056

21) Shufang, Z. (2021). Design of an Automatic English Pronunciation Error Correction System Based on Radio Magnetic Pronunciation Recording Devices. *Journal of Sensors*, *2021*. https://doi.org/10.1155/2021/5946228

22) Sun, W. (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: a mixed methods investigation. *Frontiers in Psychology*, *14*(August). https://doi.org/10.3389/fpsyg.2023.1210187

23) Thomson, R. (2011). Computer Assisted Pronunciation Training: Target- ing Second Language Vowel Perception Improves Pronunciation. *CALICO Journal*, *28*(3), 744–765.

24) Tim Schlippe, Sebastian Ochs, T. S. (2014). Web-based tools and methods for rapid pronunciation dictionary creation. *Speech Communication*. ttps://doi.org/https://doi.org/10.1016/j.specom.2013.06.015

25) Willoughby, L. (2024). *Online videos for self-directed second language learning*. *28*(1), 1–15.

26) Yang, S., & Lv, J. (2021). Adaptive Pronunciation Proofreading of Spoken English in a Wireless Sensor Network Environment. *Journal of Sensors*, *2021*. https://doi.org/10.1155/2021/2997928

27) Yolchuyeva, S., Németh, G., & Gyires-Tóth, B. (2019). Grapheme-to-phoneme conversion with convolutional neural networks. *Applied Sciences (Switzerland)*, *9*(6), 1–17. https://doi.org/10.3390/app9061143

28) Zhang, Y., & MacWhinney, B. (2023). Using diagnostic feedback to enhance the development of phonetic knowledge of an L2: a CALL design based on the unified competition model and the implementation with the Pinyin Tutor. *Language Testing in Asia*, *13*(1). https://doi.org/10.1186/s40468-023-00232-6

29) Zhou, Y., & Zhao, G. (2022). English Pronunciation Calibration Model Based on Multimodal Acoustic Sensor. *Journal of Sensors*, *2022*. https://doi.org/10.1155/2022/2208653