Enhancing User Experience through Improving the User Interface of Phonetics Tools and Studies on Phone-level ASR-based Automation through Deep Learning Techniques

by

Chang Ren

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 5, 2023

Keywords: User Experience, E-learning, Speech Disorders, Automatic Speech Recognition, Deep Learning

Copyright 2023 by Chang Ren

Approved by

Dr. Cheryl D. Seals, Chair, Professor of Computer Science and Software Engineering
Dr. Wei-Shinn Ku, Professor of Computer Science and Software Engineering
Dr. Jakita O. Thomas, Associate Professor of Computer Science and Software Engineering
Dr. Yang Zhou, Assistant Professor of Computer Science and Software Engineering
Dr. Dallin Bailey, Associate Professor, Department of Communication Disorders, Brigham Young University

Abstract

The research includes three studies at the intersection of communications disorders and computational linguistics. We begin with the case study of APTgt, a system created to improve reinforcement for Phonetics students and improve Linguistic tools for their instructions. A portion of this system utilizes machine learning techniques (i.e., Multi-class classification) to automatically generate exams. After the utilization of this learning technology, we endeavored to enhance the user experience by automatically transcribing user speech into phoneme level in research Grapheme-to-phoneme (G2P) conversion from English text to IPA format to support phonetic transcription and automatic exam generation. From the literature, we have seen support for standard speech through G2P but have found no evidence of support for disordered speech. We utilize Automatic Speech Recognition (ASR) with deep learning techniques to recognize disordered speech. This study will improve user experience and user interface design and incorporate deep learning techniques to provide phonetic transcription for disordered speech. Deep learning techniques were utilized to support the development of a Speech-to-IPA module for disordered speech and increase user efficiency by generating a large number of phonetic transcription exam resources as a word bank for exam development.

Acknowledgments

I am deeply grateful for the invaluable guidance, constant support, encouragement and experience sharing I received during my study at Auburn University. Without the assistance of my advisor, collaborators, and family members, this dissertation would not have been completed.

First and foremost, I would like to extend my sincere and deepest gratitude to my advisor, Dr. Cheryl Seals for her creative thinking, patience, unwavering support, encouragement, and dedication over the past years. The professional bond and friendship we have forged will forever hold immense value to me. I will never forget her extensive knowledge and inexhaustible enthusiasm for research, which continues to inspire and drive me toward achieving my goals.

I am also tremendously grateful for the guidance and contributions from my committee members, Dr. Wei-Shinn Ku, Dr. Jakita O. Thomas, Dr. Yang Zhou, and Dr. Dallin Bailey. Their careful review of my proposal and dissertation documents, along with their valuable suggestions, have significantly enhanced the quality of my research. I would also like to express my appreciation to Dr. Angela I. Calderón for serving as my university reader.

In addition, I owe my gratitude to Jueting Liu, Yaoxuan Luan, Yang Cao, Onyinye Asogwa, and Majdi Lusta whose assistance with research ideas, data collection, paper writing, and engaging group discussions have been invaluable. I am also grateful to my friends Muzi Li, Jingjing Li, and Boning Liang, as well as all the professors and students in the Department of Computer Science and Software Engineering. The excellent study and research atmosphere created and maintained by everyone has contributed to my growth.

Finally and most importantly, the boundless love from my family serves as my ultimate source of strength in my research journey. My mother Huimin Li, my father Hongxian Ren,

My mate Dongji Feng, and all other family members have consistently cheered me on and supported me through every challenge. I am incredibly blessed to have my dear little one, Luca Feng, who brings immense joy to my life, fueling me with the energy and determination needed to surmount obstacles. Their steadfast presence and belief in my abilities have been vital to my success.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

APTgt Automated Phonetic Transcription Grading Tool

ASR Automatic Speech Recognition

CMDS Communication Disorders

CRNN Convolutional Recurrent Nerual Network

G2P Grapheme-to-Phoneme

IPA International Phonetic Alphabet

MFCCs Mel-frequency Cepstral Coefficients

PER Phoneme Error Rate/Phone Error Rate

UCD User Centered Design

UID User Interface Design

WER Word Error Rate

Chapter 1

Overall Introduction

This research focuses on enhancing the Automatic Phonetic Grading Tool (APTgt), an online educational content management system designed to support instructors in communication disorders. Improving the tool is achieved through the development of an improved user interface. This work also aims to create a framework and deep-learning tools to improve the pedagogical experience of linguistic instructors with efficient tools for phonetic transcription, training, and exam generation.

In Study 1, the research focuses on optimizing the user interface of APTgt. By incorporating user-centered design principles and feedback from communication disorder specialists, the goal is to improve ease of use, aesthetics, and consistency in design. The study aims to reduce the number of actions required, provide visual cues for easier navigation, and enhance the overall usability of the phonetic E-learning system.

Study 2 addresses the need for accurate phonetic transcription in the system. A neural machine translation (NMT) tool is developed using the Transformer architecture to automatically translate English words into their corresponding IPA phonetic spelling formats. This tool aims to eliminate the manual input required for generating phonetic transcription exams, making the process more efficient for instructors. Additionally, it annotates speech samples in the speech corpus for automatic phone recognition, further enhancing the functionality of the system.

In Study 3, the focus is on the development of the Speech-to-IPA module within the APTgt system. This module is designed to convert speech directly into IPA-based phonetic transcriptions using Automatic Speech Recognition (ASR) techniques. By employing Mel-frequency cepstral coefficients (MFCCs) as features and utilizing a bidirectional Long Short-Term Memory (LSTM) model architecture, the Speech-to-IPA module enables instructors to automatically generate phonetic transcriptions for not only normal speech but also the disordered speech. The module leverages the Speech Exemplar and Evaluation Database (SEED) for training and testing on normal speech data. Additionally, the TORGO Dysarthric Speech Database is utilized for training and testing on disordered speech data. This integration enables instructors to generate a diverse range of phonetic transcription exam resources automatically, eliminating the need for manual transcription. This advancement enhances the intelligence and functionality of the phonetic E-learning system, benefiting instructors in resource creation and supporting auto exam generation.

This research aims to improve the APTgt system by enhancing the user interface, developing a G2P conversation tool, and implementing a Speech-to-IPA system. The integration of machine learning technology to reinforce the phonetic tool used by linguistic instructors to automatically generate and grade phonetic exams. This automation makes research in communication disorders more tractable, as it allows for the transcription of disordered speech. These advancements will provide linguistic instructors with more efficient tools for phonetic transcription training and exam generation, bring diverse speech samples into the classroom and offer more practice opportunities for students in the field of communication disorders, ultimately improving their pedagogical experiences and the overall effectiveness of the phonetic E-learning system.

Chapter 2

Optimization on APTgt system user interface

2.1 Introduction

2.1.1 Background

As technology has been widely adopted by younger generations and has become a primary necessity in university classrooms over the past decade, the demand for e-learning has increased significantly in recent years[27, 44]. In the field of communication disorders (CMDS), the clinical phonetic transcription skill is a critical part of linguistic undergraduate/graduate students' clinical preparation to become speech-language pathologists. However, they often report feeling unprepared to apply the skill in clinical practice as the practice opportunities can be impeded by the limited resources for linguistic instructors to manage the grading of additional assignments through traditional learning approaches[48]. Therefore, we began to investigate this case study, i.e. creating a web-based, integrated, interactive phonetics E-learning system as design and development challenges with opportunities to provide a rich learning experience[44]. The Automated Phonetic Transcription Grading Tool (APTgt) was developed in our HCI lab to conduct online course content and automated grading of transcription assignments.

The APTgt is an interactive and engaging online educational content management system designed to support communication disorders faculty with reinforced linguistic transcription and specialized services not currently supported by other content management systems[44]. This platform offers a convenient mechanism to support instructors in creating and managing online courses and resources. The traditional method of attending phonetic exam require linguistic students to phonetically transcribe works presented via audio recording and the traditional method of grading phonetic transcription exams involves instructors manually evaluating each student's transcriptions of the recordings based on accuracy and completeness. This can be a time-consuming and arduous task, leading to limited opportunities for students to practice their transcription skills. In contrast to the traditional method, APTgt provides an automated grading system that saves time and reduces the burden of grading and providing feedback for instructors (See Fig. 2.1).

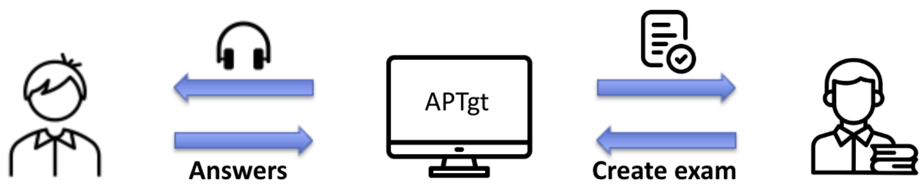


Figure 2.1: Workflow of Phonetic Exam in APTgt

While it shares some fundamental functionalities with other learning platforms, such as delivering course content and sharing materials, APTgt stands out with its unique features, including the incorporation of Embedded International Phonetic Alphabet (IPA) Keyboards[4, 35] (See Fig. 2.2). During the exam, teachers upload audio files to create exam questions, which students listen to and answer using the IPA keyboard. The grading module automatically generates a grade by calculating the similarity between the submitted answer and the pre-stored correct answer using the Levenshtein distance algorithm. This automated system allows instructors to provide real-time practice and reinforcement of students' transcription skills through immediate feedback. Upon thorough investigation, no analogous system was discovered, thus establishing APTgt as a highly innovative and interactive E-Learning platform[35, 31].

2.1.2 Research Problem & Motivation

The implementation of APTgt involved the use of a range of front-end and back-end web development techniques. While our previous research mainly focused on the functions,

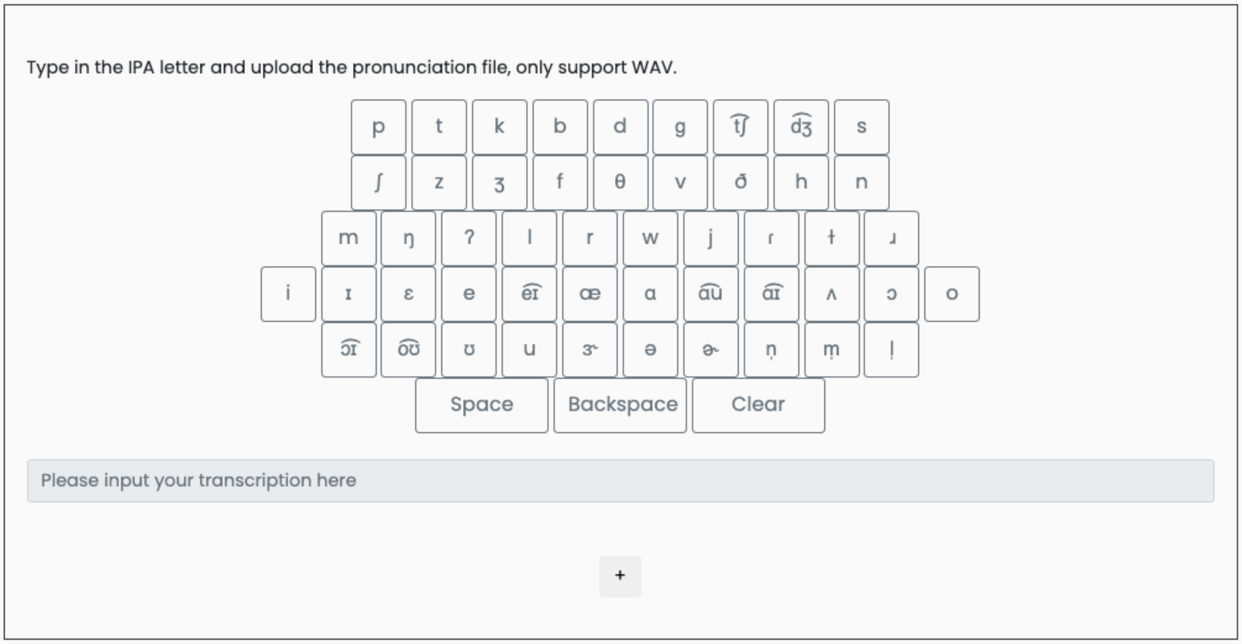


Figure 2.2: APTgt IPA Keyboard[31]

algorithms, and back-end technologies of the platform, we have identified areas for improvement in the front-end user interface design and user experience. Some common issues we have identified include cluttered layouts, lack of visual hierarchy, inconsistent design, and unintuitive navigation. These problems can lead to user frustration and reduced productivity when using the platform[44, 31].

In this research, we will focus on the next iteration of the design process and address user interface design and user experience concerns for linguistic instructors. This study proposes the optimization of the user interface of this interactive system to improve ease of use, improve aesthetics, and emphasize a more consistent design. We will increase the efficiency by reducing the number of actions required and providing more visual cues for ease of navigation. Based on user experience from participatory design partners (i.e., communication disorder specialists), my research aims to optimize current designs to provide user-centered experiences for linguistic teachers and redesign UIs and increase the time efficiency of using E-learning software. During the iterative design process, we will follow the User Centered Design (UCD) principles and focus on the user and their needs in each phase.

2.1.3 Research Questions

Based on user experience from participatory design partners (i.e. experts in communication disorders), my research aims to address the following questions:

1. How can the user interface of APTgt be optimized to improve ease of use, aesthetics, and consistency in design?

2. How can the efficiency of APTgt be increased to reduce the time required for instructors to use the software?

2.1.4 Research Hypothesis

1. Aesthetically Pleasing:

   The layout and design of the application interface are intended to be aesthetically pleasing.

   > $H1_0$: There is no difference between the experimental and benchmark interfaces on individual appeal, streamlining of features (reduction of complexity), and organization of the information.

2. Ease of Use:

   The overall application design is intended to be perceived as easy to use.

   > $H2_0$: There is no difference between the experimental and benchmark interfaces on ease of use.

3. Satisfaction:

   The overall design of the application is intended to satisfy users.

   > $H3_0$: There is no difference between the experimental and benchmark interface on overall satisfaction.

The contributions of this study will be the improved user interface and user experience of APTgt, which will provide a more efficient and usable tool for communication disorders faculty. By addressing the user interface design and user experience concerns, this study aims to optimize the current designs to provide user-centered experiences for instructors and increase the time efficiency of using e-learning software. Additionally, the study will contribute to the field of E-learning by demonstrating the importance of user-centered design in developing effective online educational tools for instructors.

2.2 Literature Review

2.2.1 E-Learning

The article "Practice makes perfect? The pedagogic value of online independent phonetic transcription practice for speech and language therapy students[49]" describes a study performed on a cohort of students studying phonetic transcription and speech disorders. The study involved two parts: giving the students a weekly quiz (the 'Ulster Set') on a specific accent and also giving the students access to an online practice platform (WebFon). Student engagement with WebFon was measured in terms of the number of responses made to 'sparks' (weekly questions posted by the lecturer) on the University's Virtual Learning Environment Discussion Board. Measures of phonetic transcription accuracy were obtained for the 'Ulster Set' and for the final phonetic transcription coursework at the end of the module. Qualitative feedback about the experience with online learning was gathered via a questionnaire. WebFon and the 'Ulster Set' both allow students to listen to audio files and respond using the phonetic keyboard, the UCL Unicode phonetics keyboard.

Researchers found a positive correlation between 'Ulster Set' scores and usage of WebFon. In addition, there was a positive correlation between 'Ulster Set' scores and final transcription assignment scores. Students generally found WebFon to be a useful tool as well as thought that the 'Ulster Set' was good preparation for their work in the real world. Overall, the use of these tools had a positive impact on the student's coursework. The authors also note that students often struggle more with vowels than consonants. These programs and quizzes should be designed to allow students to practice the topics they need most. With software, it's possible to give more regular practice and better enable students to distinguish between accents and disorders. Furthermore, the author discusses the importance of considering the ease of use of online learning resources from a technical point of view, as any difficulties in accessing or navigating the resources can negatively impact user engagement. The article highlights that the online practice tool used in the study, WebFon, was reported by students as being easy to use, but there were some technical issues with the "Ulster Set" resource. Therefore, it is important to ensure that online learning resources are user-friendly and compatible with current IT trends and needs.

In the field of E-learning, Moodle is also a widely recognized and extensively used learning platform. It is designed to create personalized learning environments for educators, administrators, and learners[6]. To utilize Moodle, educational institutions need to download and install the software package on a web server, which can be a personal computer or a hosting service. It is compatible with various operating systems, including Windows, Mac, and Linux, and requires PHP and SQL database support. Moodle, which stands for Modular Object-Oriented Dynamic Learning Environment, caters to both programmers and educators with advanced computer skills. The term "Moodler" refers to anyone who uses Moodle, reflecting the flexible and adaptable nature of the platform[50].

2.2.2 User Interface Design

In the paper "User Interface Design for E-Learning Software[15]", the User Interface is defined by which the user and a computer system can interact. The success and failure of any software system can be correlated to the effectiveness, efficiency, and user satisfaction with the User Interface Design. Nowadays, UI is a crucial factor in designing any educational software. Principles and concepts of learning should be considered in addition to UID principles in UID for e-learning. Faghih et al. discuss the role of User Interface design in an e-learning application software. According to the User Interface Design of e-learning software, the psychology of the student or learner is an important aspect to be considered while developing an e-learning application. The UI is the point of interaction between the user and the education body, so if requirements are not implemented to support such correlation then our aims of education may not be achieved. UI designer arranges elements (such as multimedia, and tools like Textbox, Label, etc.) with which users can use the computer more easily. The design begins with an understanding of the intended users, including profiles of their age, sex, physical abilities, education, cultural or ethnic background, motivation, goals, and personality. There are 3 golden rules for designing UI mentioned in the paper: (1) Place the user in control; (2) Reduce the user's memory load; (3) Make the interface consistent. E-learning applications should be designed in such a way that the pervasive feeling of requirement and motivation grows constantly, and the coercion feeling reduces. Motivation is the key factor to be considered in e-learning systems and their growth, some suggestions to increase motivation are as follows:

- Using speech interface;
- Using an informal communication style instead of a formal;
- Using a variety of colors in educational Media;
- Using background music;
- Learners have control over the learning environment.

Availability is also an important aspect of E-learning systems, which means that users can easily access intended content. Allowing learners to access previously taught materials regularly or can look for specific content in the e-learning system anytime. Moreover, whenever the words or phrases that are used in the text exist elsewhere, they should act as a link to navigate, describe, and return the user to the previous page simply.

Overall, the paper provides a comprehensive review of user interface design for E-learning software and offers practical guidelines and recommendations for designers and developers to create effective and user-friendly interfaces in E-learning systems.

Oleksiy[26] proposed a framework for user-centric personalized UI development, utilizing configurable UI elements that adapt to user preferences. This approach aims to streamline UI development efforts and focus on improving product business logic. The framework incorporates a distributed architecture for personal semantic user-profile management, allowing users to make immediate profile changes within specific contexts. The paper also addresses the challenge of semantic API visualization and presents an approach for on-the-fly UI creation based on machine-readable semantic descriptions. The implementation of this approach in government-funded projects facilitates customer engagement and feedback gathering, leading to a better understanding of customer needs and the development of products that meet those needs. Overall, the paper emphasizes the significance of semantic personalization in feedback-supportive tools for effective customer involvement in the development process.

2.3 Methodology

2.3.1 Current system

The APTgt system is an interactive E-learning system developed by the Auburn University HCI group in a participatory design process with the faculty from the Department of Communication Disorders (CMDS) to facilitate phonetic transcription training for CMDS students [44, 48]. It is currently an online system including a well-built database schema and server environment. APTgt provides many supportive features such as the following: phonetic course content, lessons in the form of videos, practice sessions, and exam sessions. There are three partitions: admin, teacher, and student. Through this system, administrators can grant teachers and students access and manage registered users. Teachers can easily upload course materials, generate and manage the course, and grade exams online. Students will no longer be burdened with the need to remember complex phonetic symbols and can complete exams and practices with a better user experience. This is due to the system's inclusion of IPA keyboards, which enable users to select the required characters from the keyboards instead of having to recall them from memory [31, 45].

The following figures (see Figs. 2.3–2.5) simply illustrate the functions of the APTgt system.

Administrators can manage users and semesters stored in the database (See Fig. 2.3).

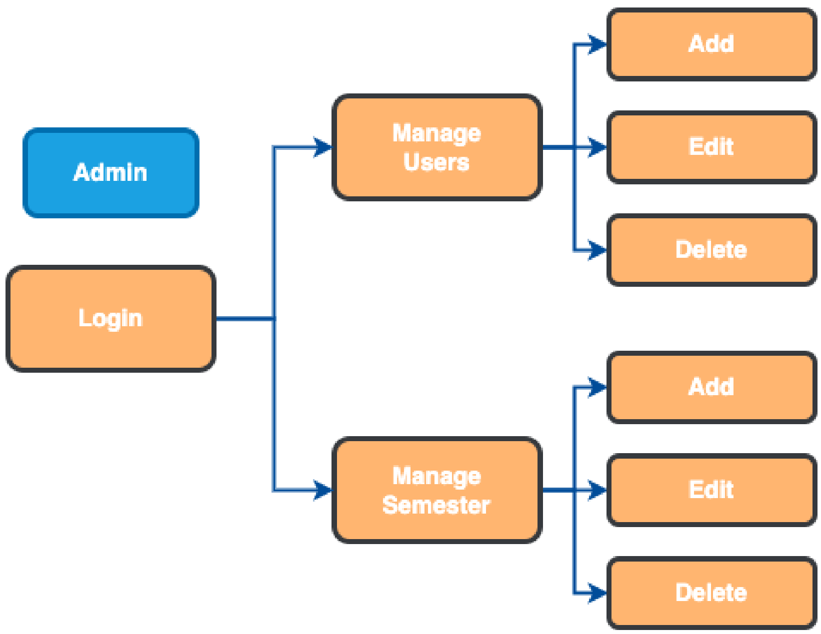


Figure 2.3: Admin view functional flow diagram

On the teacher's part (see Fig. 2.4), each teacher has his/her personal account which leads to his/her courses. The different accounts will lead to separate spaces, which means teachers cannot share lesson materials and students' information. This aims to protect students' privacy. After logging into the system, teachers can create/manage lessons, exams, and practices, and review the answers submitted by students. The lessons are formed by videos and exams are formed by audio [44, 35].

A student will have to register to the system and access the functions. Students need to enroll in some lessons to access the lesson materials and exams related to phonetic transcription. During the exam, each question is a word of pronunciation. The student needs to assemble the phonetic characters from the IPA keyboard to generate an answer that best

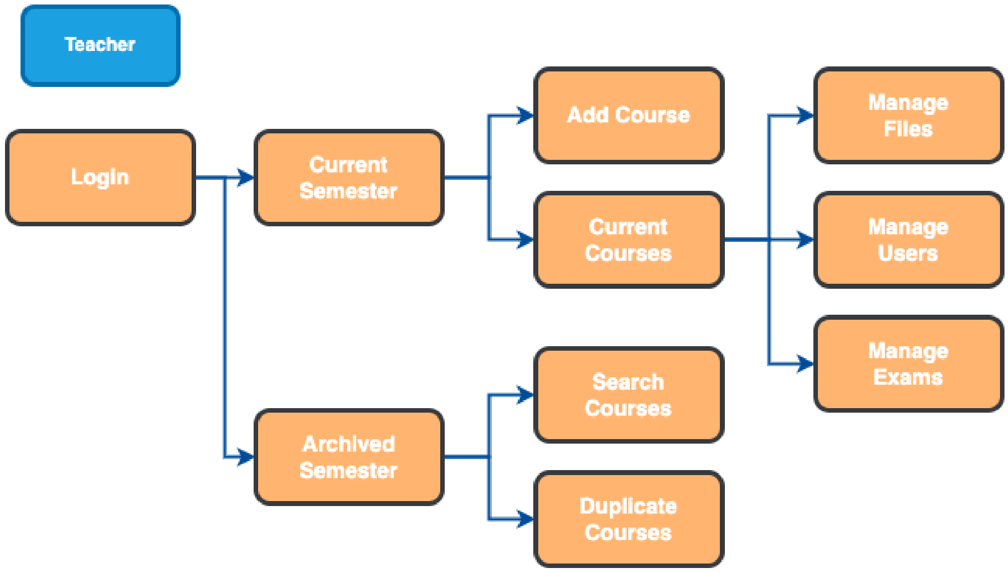


Figure 2.4: Teacher view functional flow diagram

represents the speech sound from the question. The tool also provides solutions and results analysis from students' exams, and the students will know how their grades distribute in the overall grading pole[35] (See Fig. 2.5).

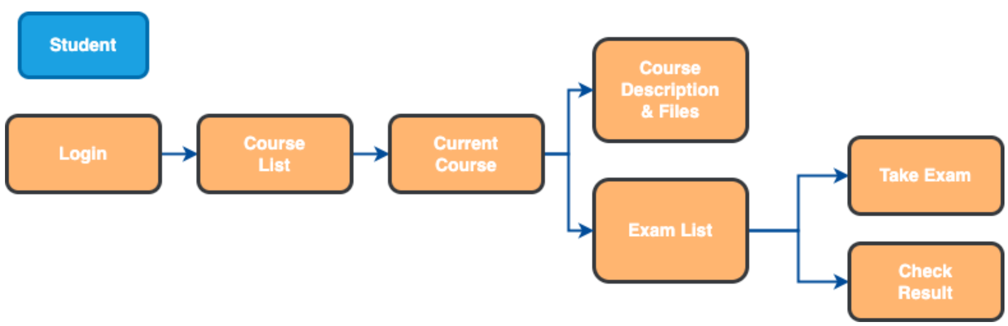


Figure 2.5: Student view functional flow diagram

The initial version of the web application prototype, encompassing functionality for users across three major roles, is showcased in Figures 2.6–2.9.

Subsequently, leveraging the wireframes as a foundation, the graphical user interface of the first iteration of APTgt (APTgt v1.0) was designed and developed by our HCI team, as illustrated in Figures 2.10–2.16.

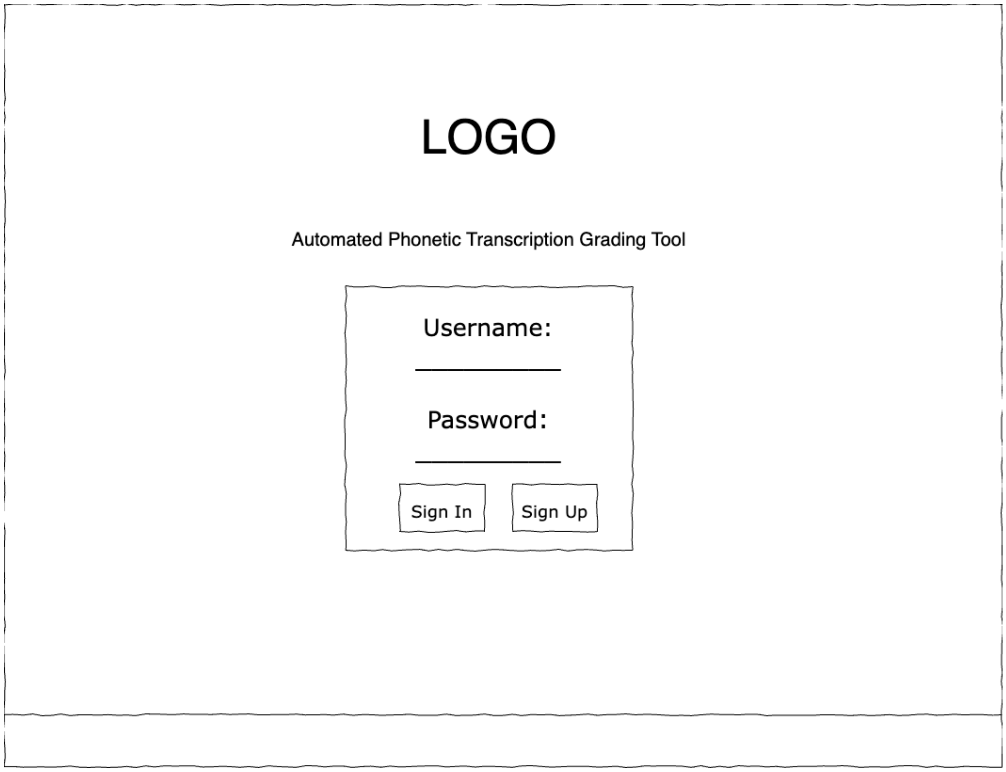


Figure 2.6: Wireframe of Sign In/Sign up Page for APTgt v1.0

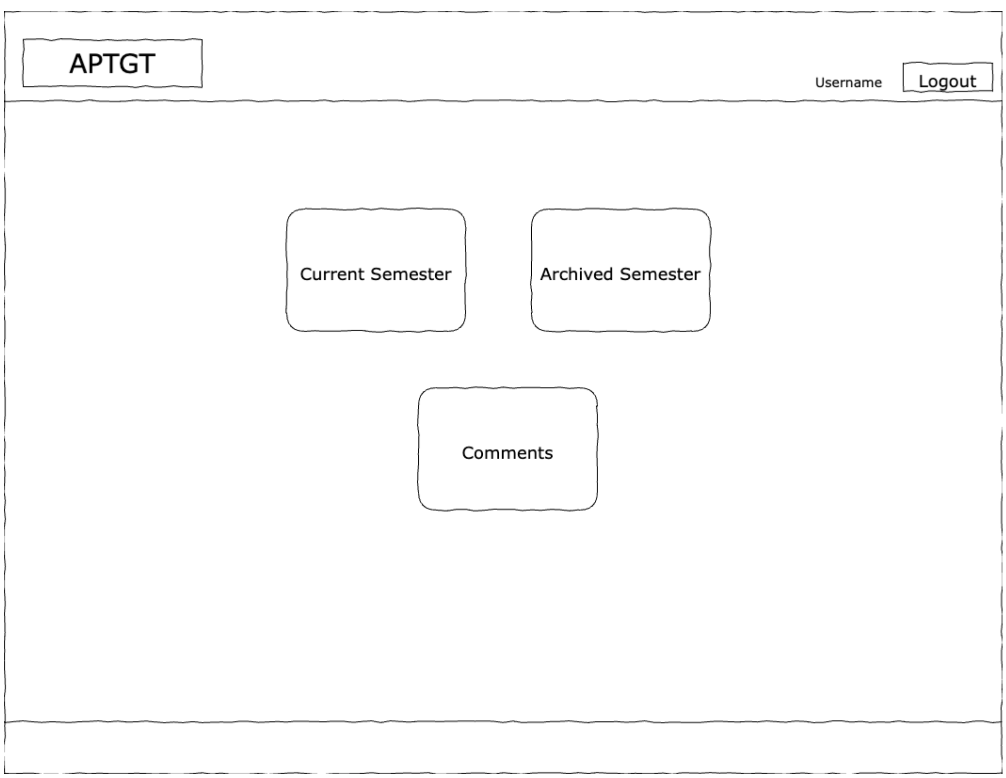


Figure 2.7: Wireframe of Homepage for APTgt v1.0

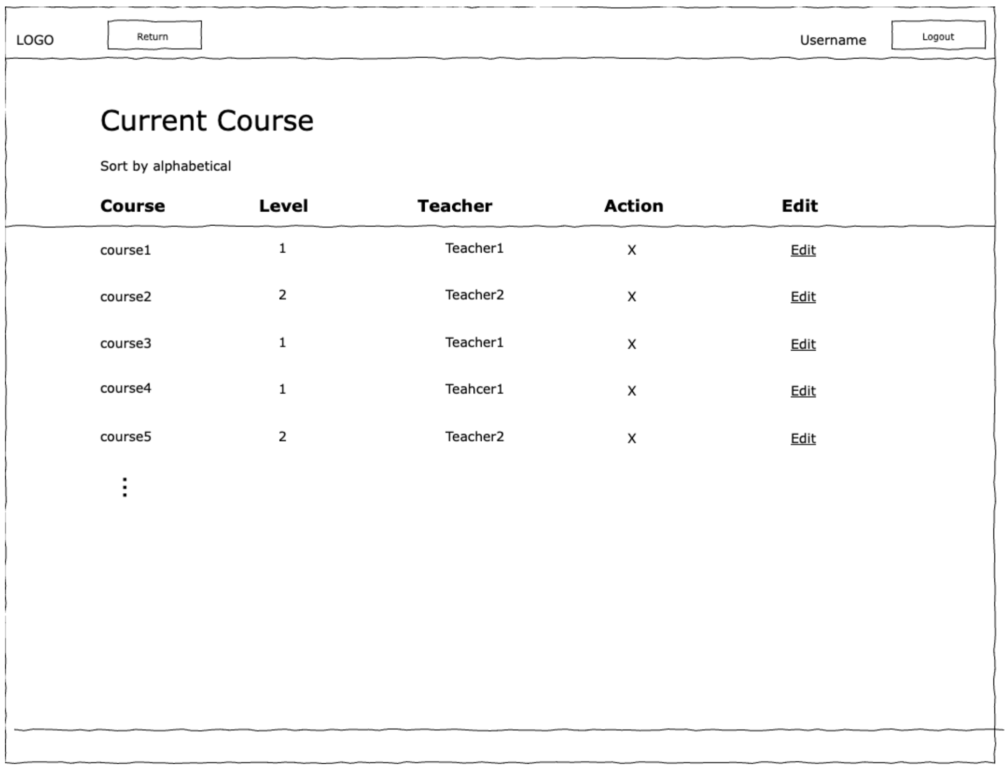


Figure 2.8: Wireframe of Current Course Page for APTgt v1.0

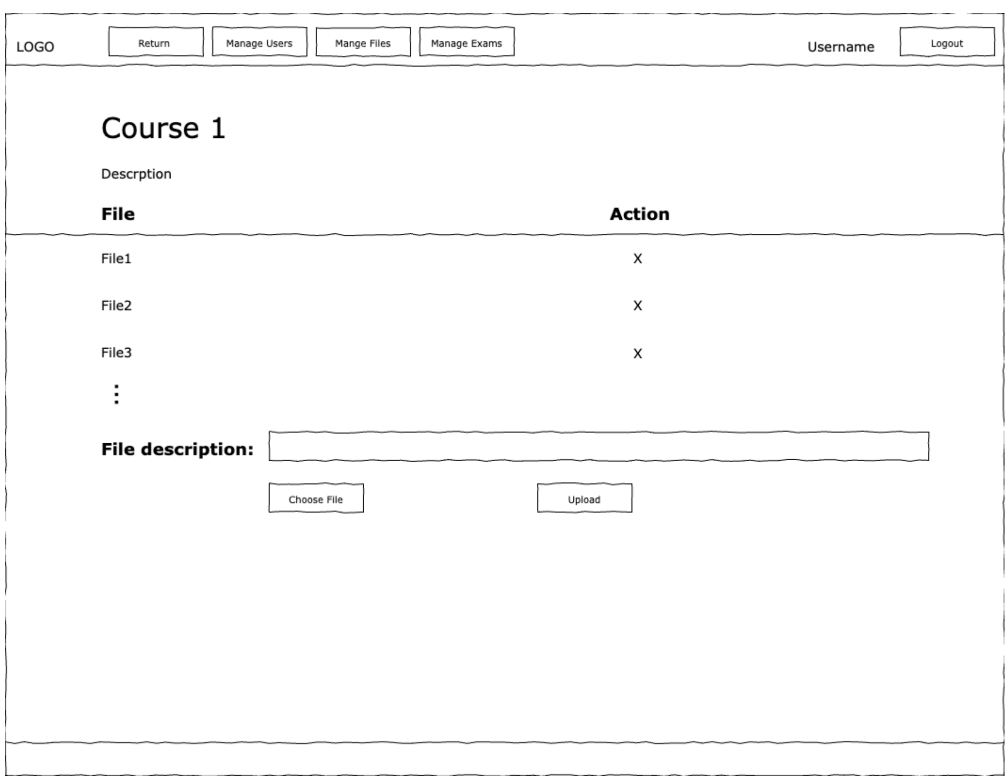


Figure 2.9: Wireframe of Course Detail Page for APTgt v1.0

Figure 2.10: Sign In/Sign up Page in APTgt v1.0

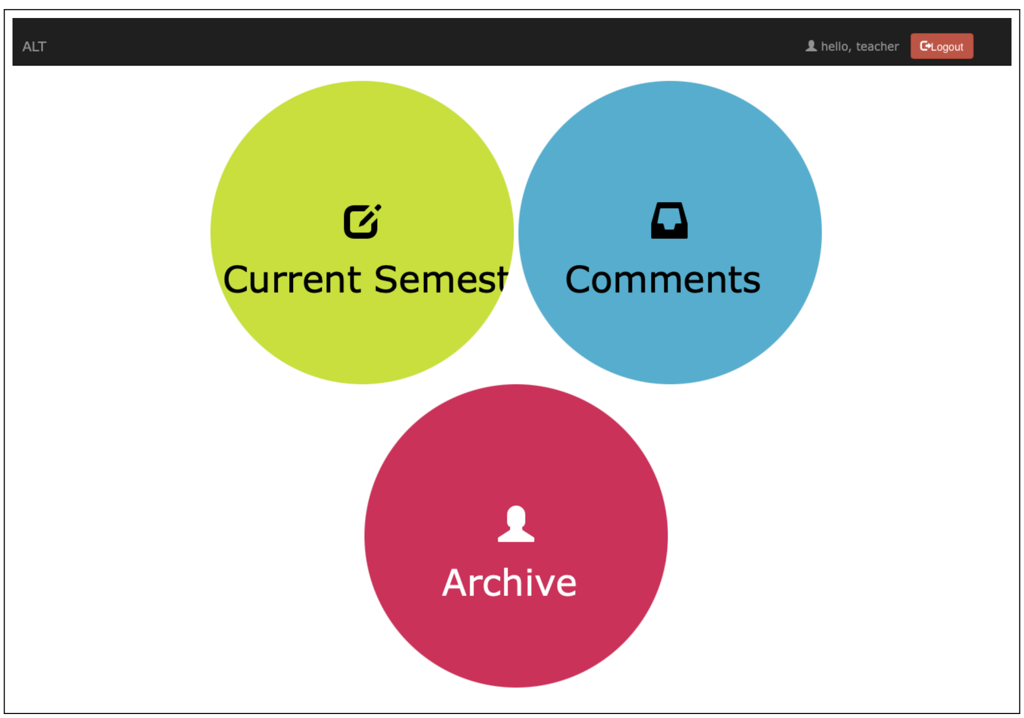


Figure 2.11: Homepage - Teacher View in APTgt v1.0

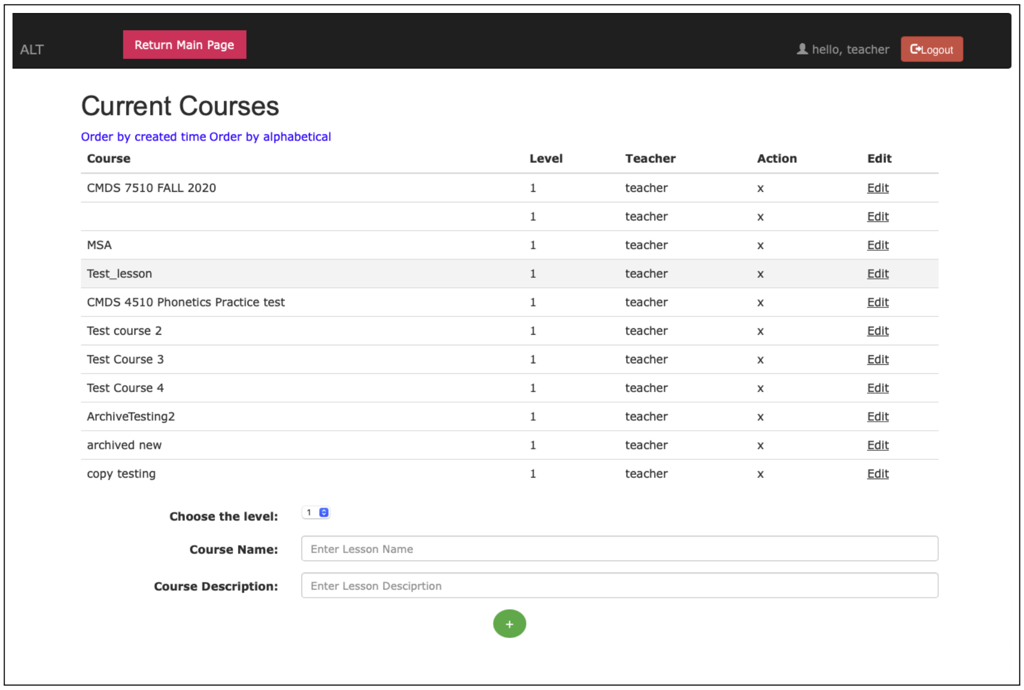


Figure 2.12: Current Courses - Teacher View in APTgt v1.0

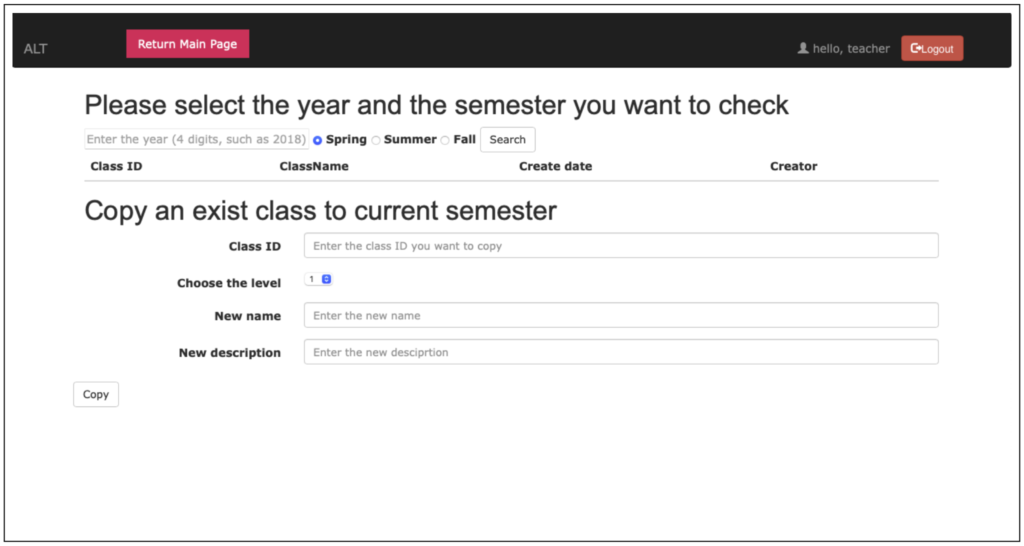


Figure 2.13: Archived Courses - Teacher View in APTgt v1.0

Figure 2.14: Homepage - Student View in APTgt v1.0

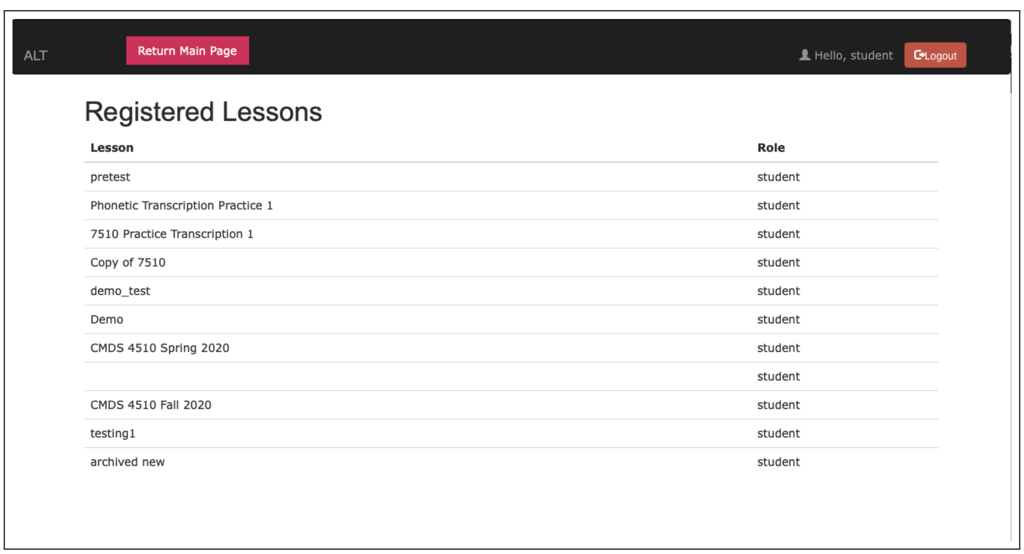


Figure 2.15: Registered Courses - Student View in APTgt v1.0

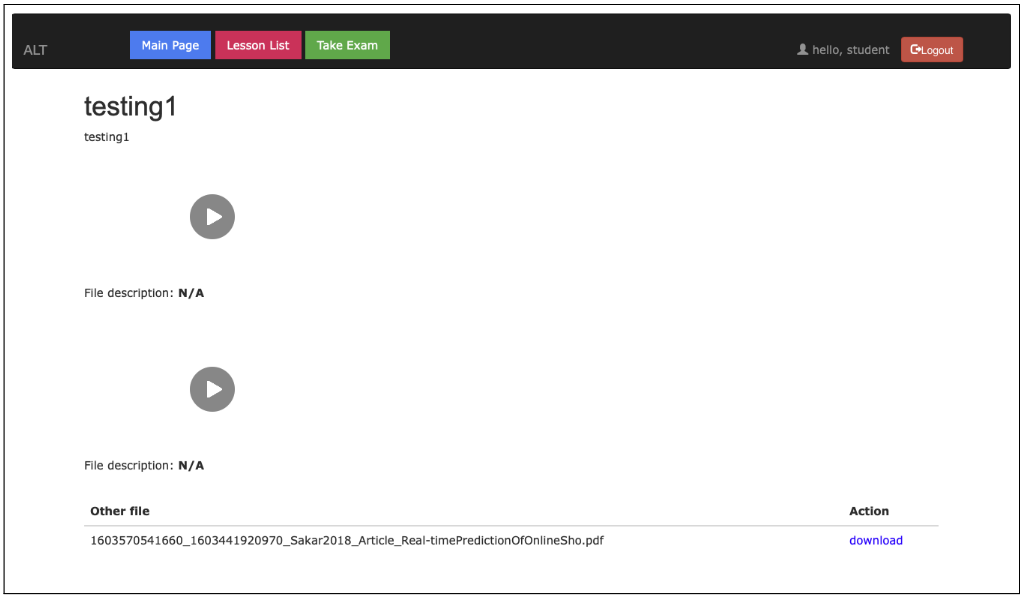


Figure 2.16: Course Detail - Student View in APTgt v1.0

2.3.2 Problem

The initial research on APTgt has focused on functionality, algorithms, and back-end technologies. However, there are notable areas within the front end that require improvement, as they can potentially result in user frustration and reduced productivity. Consequently, the team's next imperative step should concentrate on enhancing the user interface design.

Several issues have been identified in the current system, including a cluttered layout, lack of visual hierarchy, inconsistent design, poor use of color, and unintuitive navigation. For instance, the home and archive course pages within the teacher's section (see Fig. 2.11 and 2.13) possess a simple top-down layout without any division, with all text rendered in the same font color. This impedes users from swiftly scanning and locating relevant information. Furthermore, the current course and course detail pages (see Fig. 2.12 and 2.16) exhibit inconsistent design elements and poor use of color, such as the presence of differently colored buttons that can be distracting for users. Moreover, the top navigation buttons in the current course page offer limited functionality compared to a comprehensive navigation menu, as they do not provide access to sub-pages, thereby compromising usability.

To address these concerns, we propose optimizing the user interface of the interactive system by creating an easy-to-use, more aesthetically pleasing, and consistent design. We will increase the efficiency by reducing the number of actions required and providing more visual cues for ease of navigation. Despite the existence of many interconnected and integrated functions within the current system, we want an all-new design system compiling on the current server environment to improve the user experience.

2.3.3 System Requirement

The APTgt is an iterative product partitioned into three platforms: (1) Administrator, (2) Teacher, and (3) Student. The functions of each partition are described as follows:

Admin: At the administrator level, access to newly registered users is granted. And other administrative functions, such as assigning user roles (student or teacher) and deleting/disabling users if necessary.

Teacher: The teacher level of APTgt is primarily used for creating courses, embedding courses within courses, and deploying practice assignments and exams related to courses available to students. Registered teachers can create, duplicate, and edit course content for the current and subsequent semesters using their username and password.

Student: Students enrolled in the course by teachers can access uploaded content, complete practice assignments, and take exams by logging into the APTgt. Students can view their assigned course(s), assignment, due dates, and assessment scores[48].

Based on the functional requirements, the new release should meet the following criteria:

Include the following generic features for all roles:

- Login/Logout/Signup
- Reset password
- Display course content
- View exam content

- View exam results

Include the following specific features for the teacher role:

- Manage course files
- Manage Exams
- Manage Users (students)

Include the following specific features for the admin role:

- Manage Users (teachers and students)
- Manage Semesters

manage users.

In addition to the aforementioned mandatory features accessible in the new design, it requires a redesign of all web pages to have a consistent and aesthetically pleasing theme, and the layout will be implemented by providing appropriate user controls and navigation components to improve the flexibility and efficiency of use.

2.3.4 Iterative Design & Prototyping

In this study, the Scrum framework is utilized throughout the software development process, which combines the iterative design method with the incremental model(IM)[8, 16]. Figure 2.17 illustrates the iterative process of each IM cycle. As part of the IM, each increment passes through four phases: analysis, design, implementation, and testing. Every new release after each cycle adds functionality to the previous cycle.

1. Analysis Phase

The IM cycle begins with the analysis phase. At this stage, we should follow the process below: 1) Understand how the current system works, users, and system requirements;

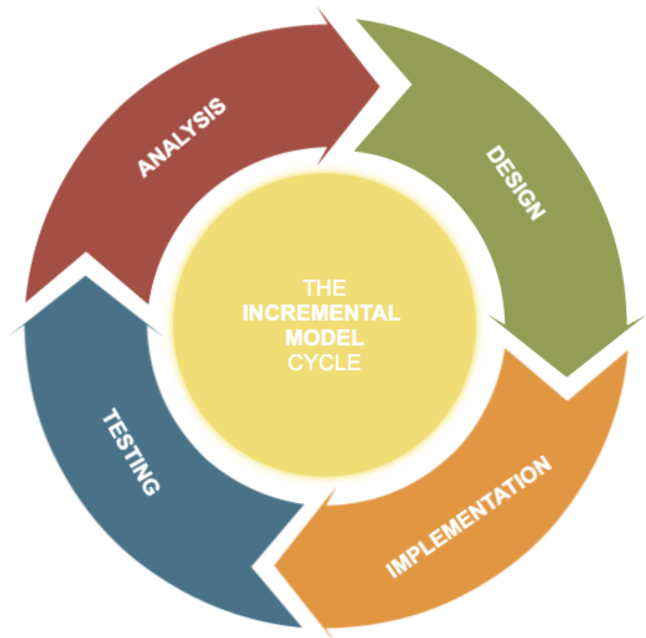


Figure 2.17: The Incremental Model Cycle

2) Gather functional and non-functional requirements; 3) Create detailed specification documents of technical requirements and review them with users;4) Analyze user comments and feedback after each review session.

2. Design Phase

In the Design phase, we need to do the following steps: 1) Define the components and architecture of a system to satisfy software engineering requirements; 2) Design solutions to problems via organizing various components into an architecture; 3) Design frontend prototypes based on previous analysis; 4) Create demo frontend pages for review.

3. Implementation Phase

In the Implementation phase, we develop the system in the following ways: 1) Build a development environment and a local database; 2) Build and improve the local database according to requirements; 3) Build components such as logo images, sidebars, and theme colors; 4) Develop components individually and combine them in each page. 5) Compile and debug the new code to make sure it works properly.

4. Testing Phase

In the Testing Phase, we perform the testing by 1) Running functional tests on each webpage; 2) Fixing issues, and running testing again. At the end of every iteration, integrate the current product into the running system, a product increment is delivered.

2.4 Implementation

This section describes how the new interface will be designed and implemented. Since the framework will be used is Scrum, the development process will be divided into several sprints. Scrum is an agile software development framework that manages software development iteratively and incrementally[8]. The Scrum process enables developers to accommodate quickly, adaptably, and flexibly to changes in requirements, making it particularly valuable in scenarios like the development of new software products where requirements may be unclear or anticipated to evolve over time.

2.4.1 Sprint 1

During Sprint 1, the analysis of the system was conducted using the old design as depicted in Figures 2.10–2.16. Simultaneously, the initial system requirements were collected from the clients, Dr. Marisha Speights Atkins and Dr. Dallin Bailey, who are professors in the Department of Communication Disorders. These requirements are outlined in Section 2.3.3. Subsequently, wireframes of the updated UI design were constructed to establish a clear hierarchy of information on pages and solidify the interface structure based on the requirements.

The significant change in the updated version of the user interface design is the transition from top-button navigation to a collapsed sidebar. This modification allows for easy access to various sections and features of the system and provides more efficient use of screen space. The sidebar remains expanded by default, allowing users to have a comprehensive view of the available navigation options. However, users have the flexibility to collapse the sidebar if desired, which can provide a better layout and help them focus on the content within the main workspace. This change optimizes the user workflow by providing a more streamlined and accessible navigation structure.

The wireframes are displayed as follows (see Figs. 2.18–2.22). In the new login page(see Fig. 2.18), the placeholder for login credentials was strategically positioned on the left side. The main page on the right prominently featured an introduction to the tool, providing new users with a comprehensive overview and impression of its purpose and functionality, allowing users to quickly understand the tool's purpose. Figure 2.19 illustrates the updated

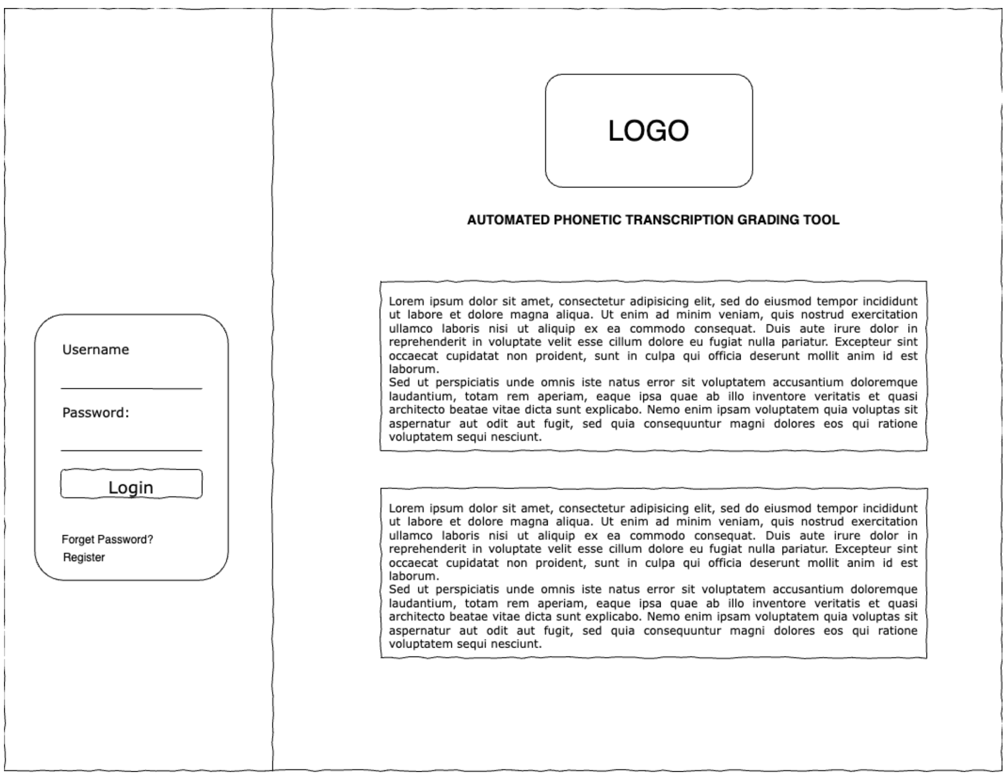


Figure 2.18: Wireframe of Sign In/Sign up Page for APTgt v2.0

home page design for the teacher role. An improvement in the new design is the inclusion of a sidebar with two distinct menus: one for the current semester and another for archived semesters. This modification reduces the number of actions required to switch between these two subpages. Unlike the previous design(see Fig. 2.11), where users had to navigate back to the main page and select a different option, the new design allows for seamless and efficient switching between the current and archived semesters.

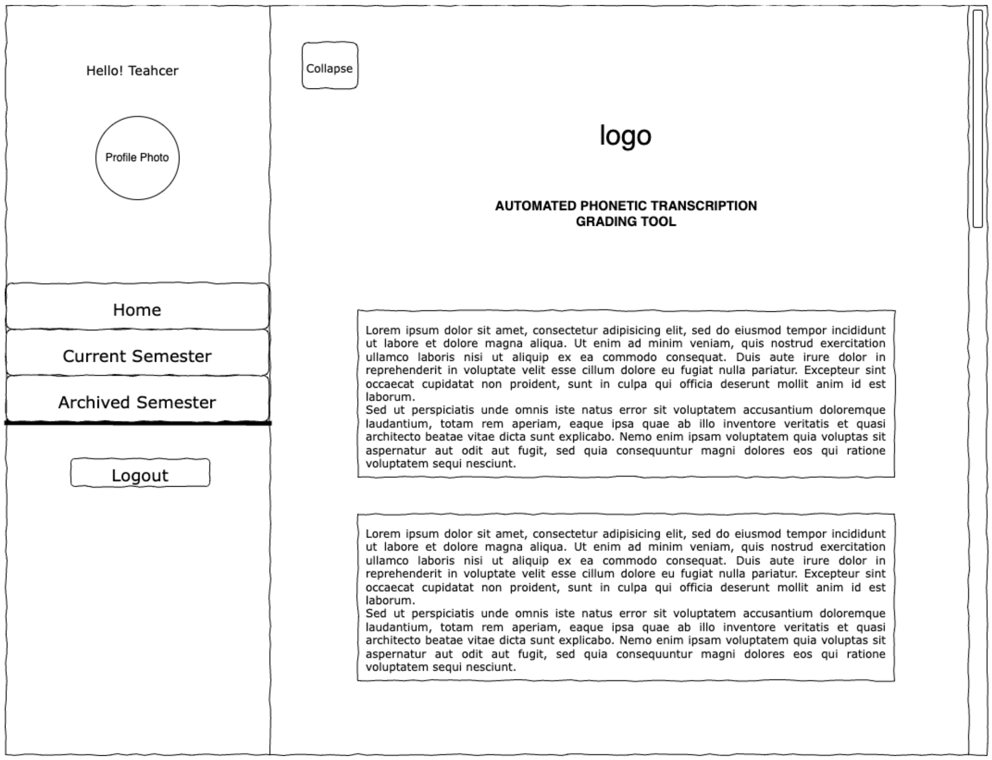


Figure 2.19: Wireframe of Home Page for APTgt v2.0

The layout of the Current Course page has been updated to provide a clear division between different sections, such as the current course and the option to add a new course(see Fig. 2.20). When a teacher selects a specific course, additional options are displayed in the sidebar, allowing them to efficiently manage class files, exams, and users related to that particular course(see Fig. 2.21). Similarly, one more option appeared for taking exams when a student selects a registered course(see Fig. 2.22). These enhancements ensure a more organized and user-friendly interface, enabling users to navigate and access the relevant functionalities with ease.

The wireframes played a crucial role in determining the placement of various components such as navigation elements, text, and image units. They also provided a clear visualization of the new layout and hierarchy, serving as a valuable reference for discussions with the clients during the next sprint.

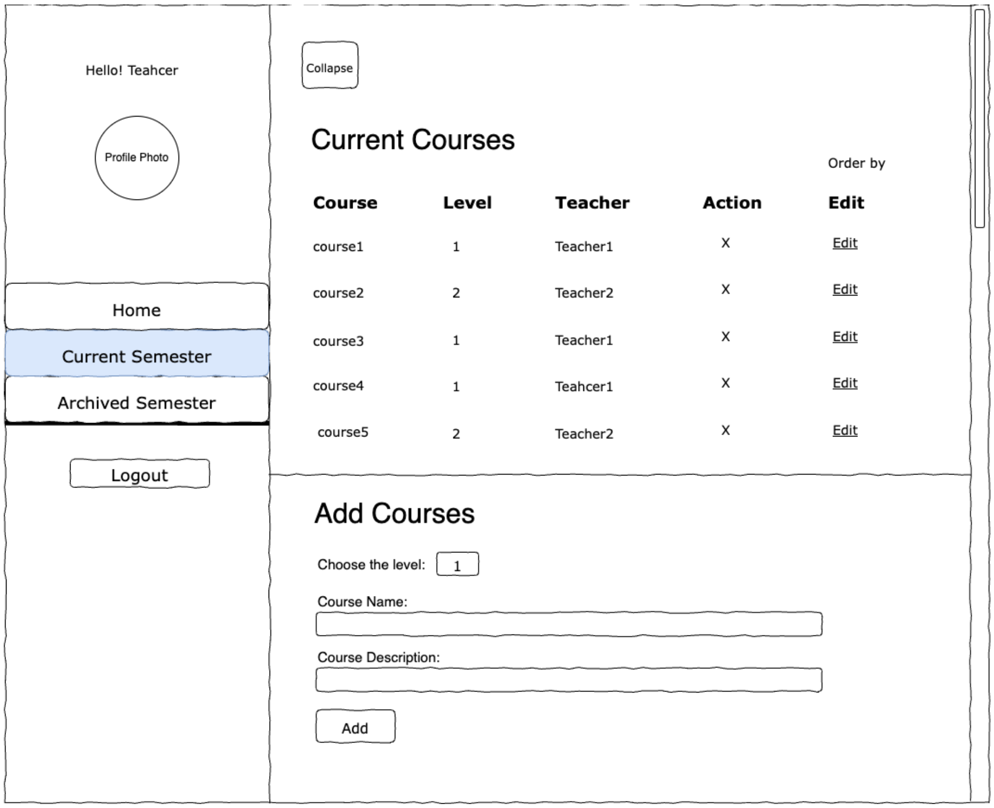


Figure 2.20: Wireframe of Current Course Page for APTgt v2.0

2.4.2 Sprint 2

During Sprint 2, the wireframes of the user interface were presented to the clients Dr. Speights and Dr. Bailey for review and feedback. Taking their input into consideration, color selections were made to align with Auburn University's branding colors, blue and orange. By incorporating these colors, the interface can achieve a cohesive and unified look, enhancing the overall user experience and ensuring a visually appealing design.

Given the requirements and wireframes, the software and hardware requirements were decided at this stage. The software requirements for a user to take advantage of APTgt are a modern web browser such as Chrome, Safari, and Firefox. As for hardware requirements, the user requires a desktop computer, laptop, tablet or smartphone, and broadband high-speed network to access the web application.

Then with a consensus reached, the team began the development phase to enhance and implement the improved interface accordingly. In keeping with the current system,

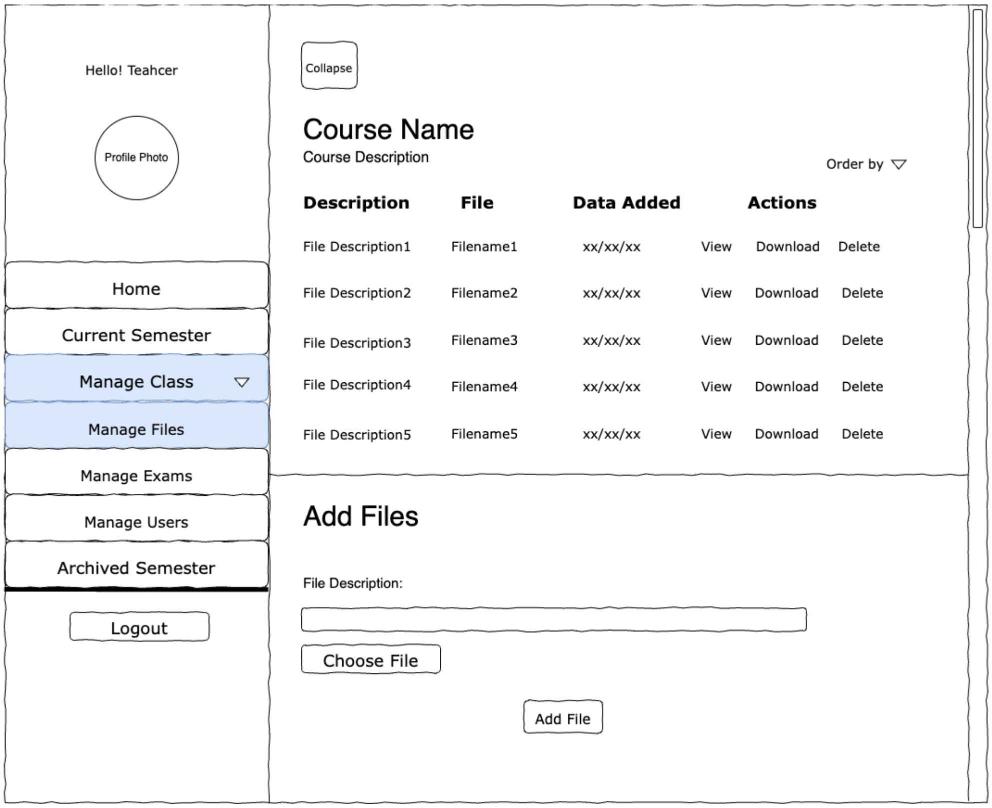


Figure 2.21: Wireframe of Course Detail Page for APTgt v2.0

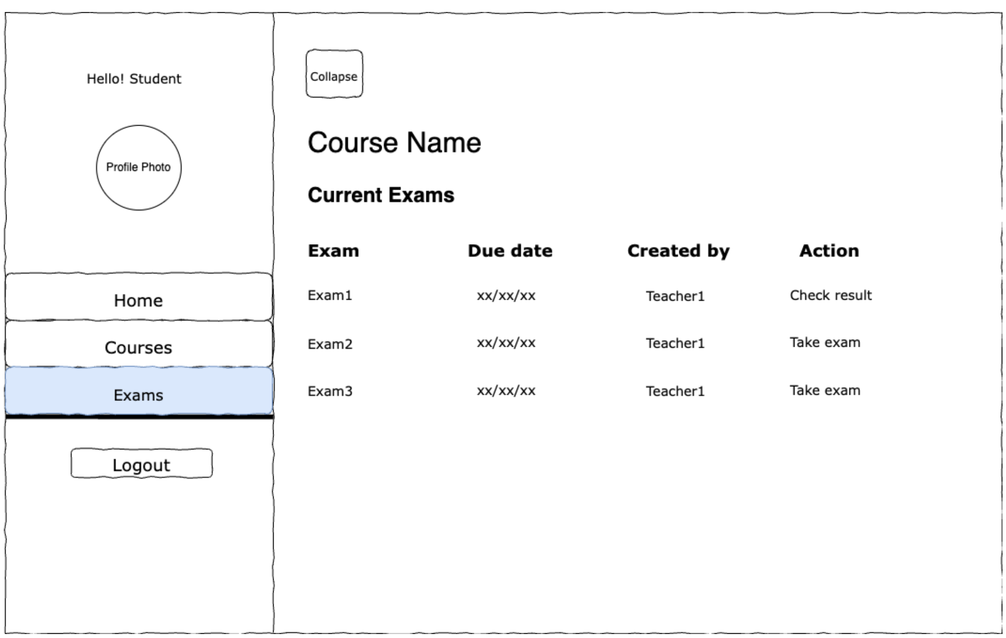


Figure 2.22: Wireframe of Exams Page for APTgt v2.0

the minimum development tools and techniques required by the development team include Windows or Mac OS, Eclipse IDE, MySQL Relational Database, and Tomcat JAVA EE Application Server. The languages and frameworks used for developing the new UI were HTML, CSS, JavaScript, and Bootstrap. Figures 2.23–2.29 showcase the updated interface and design improvements of APTgt(i.e. APTgt v2.0).

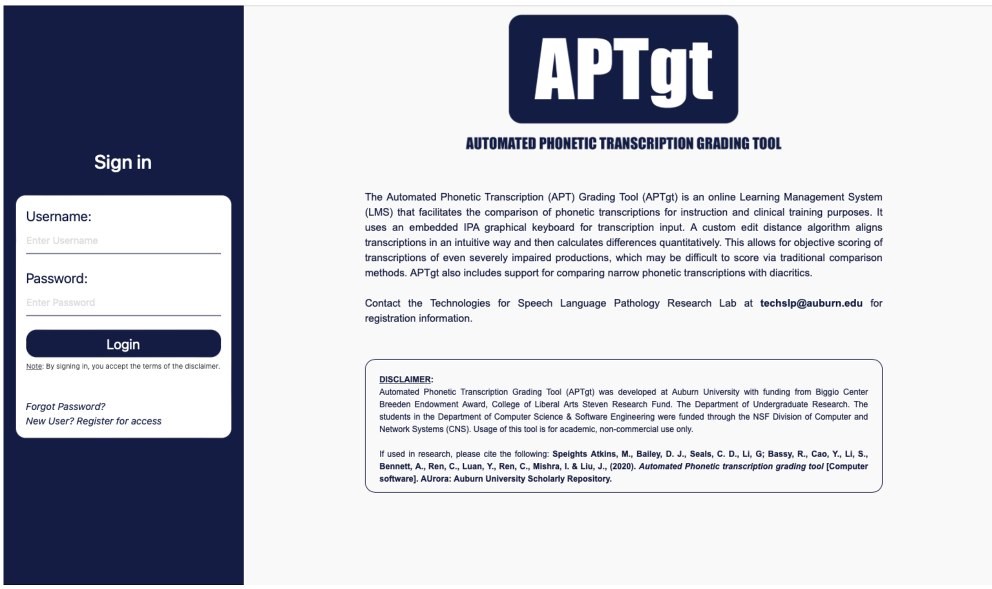


Figure 2.23: Sign In/Sign Up Page in APTgt v2.0

2.4.3 Sprint 3

The focus of Sprint 3 was to ensure all of the client's requirements were met in the new version of APTgt, and that all the functionalities were tested after integration. The test is performed in Katalon Studio which is software built on top of the open-source automation frameworks Selenium, and Appium with a specialized IDE interface for Web, API, mobile, and desktop application testing[5].

2.4.4 Testing

The testing phase was conducted in collaboration with Yuanxuan Luan and Simin Liu. In order to thoroughly test the newly designed version of APTgt, we developed a set of test

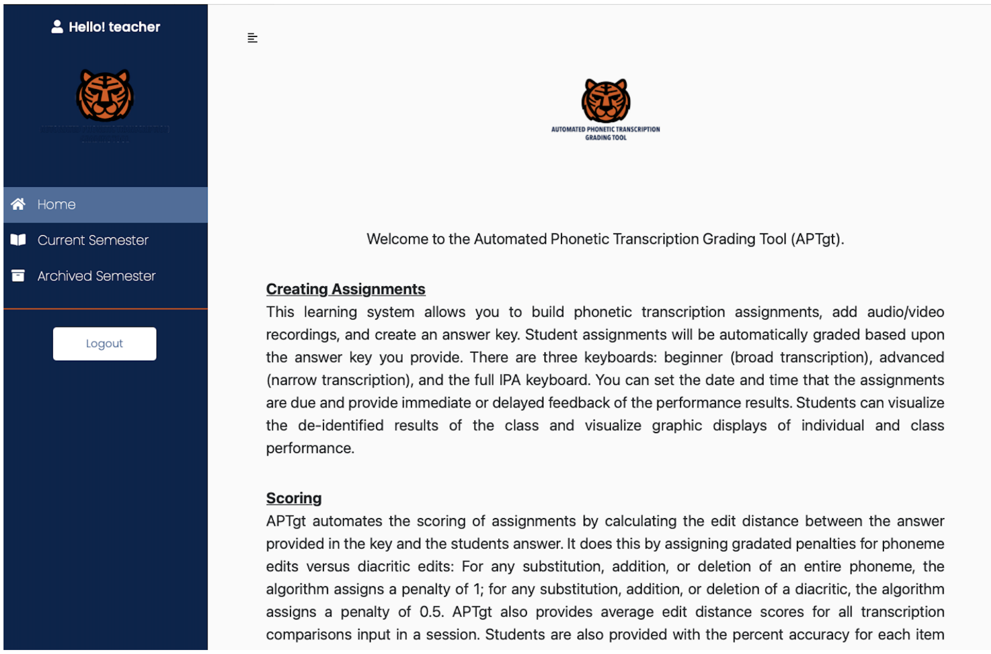


Figure 2.24: Home Page - Teacher View in APTgt v2.0

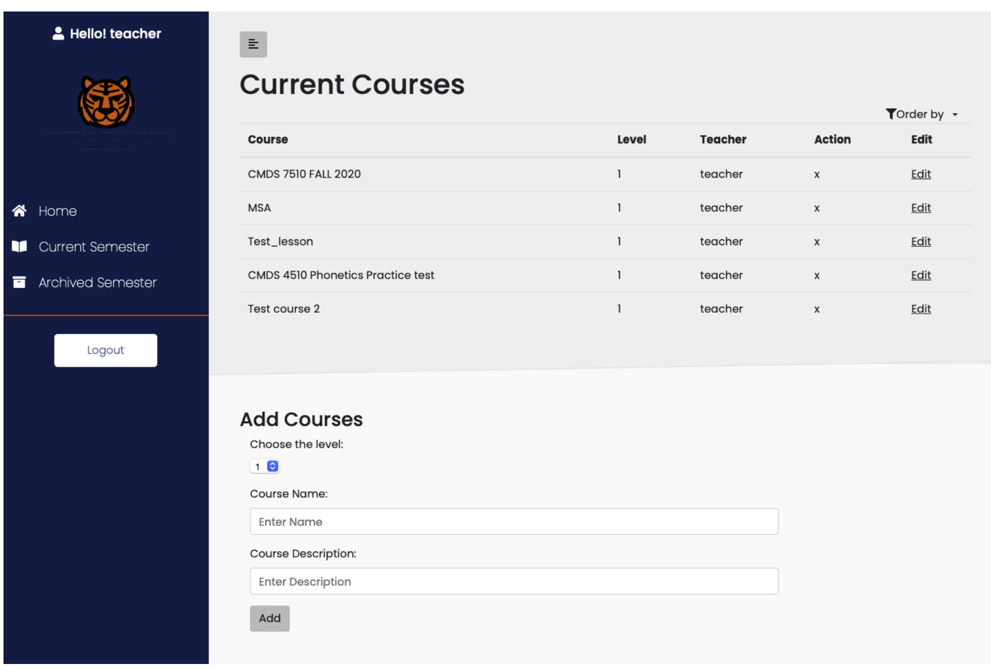


Figure 2.25: Current Course - Teacher View in APTgt v2.0

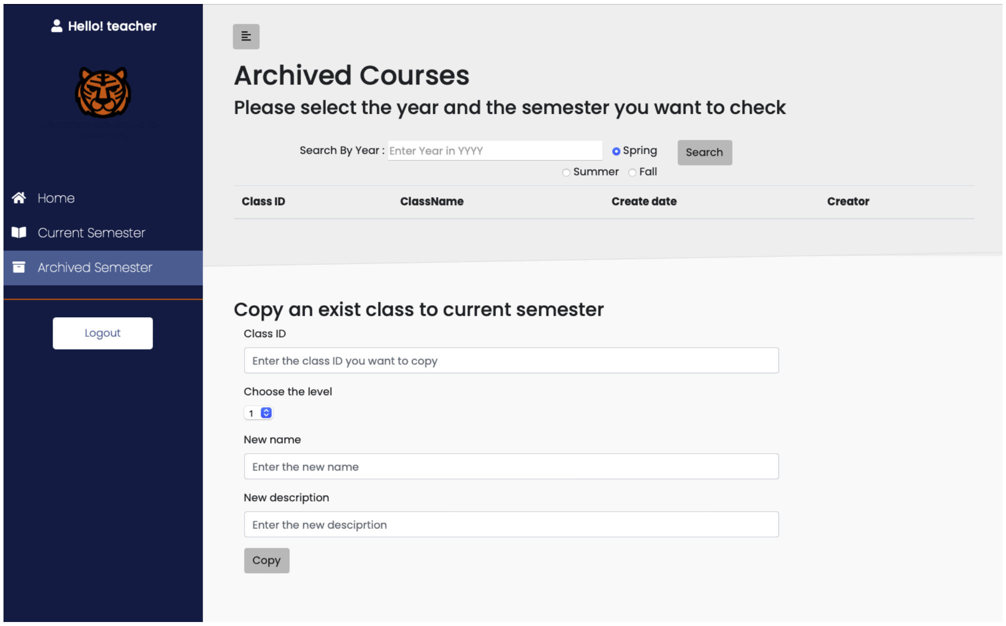


Figure 2.26: Archived Course - Teacher View in APTgt v2.0

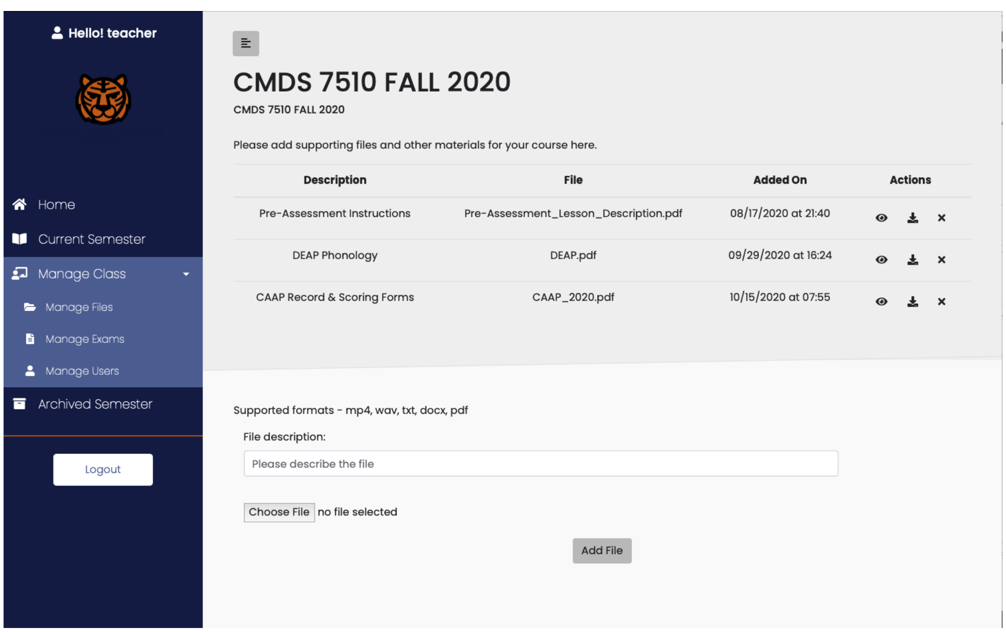


Figure 2.27: Manage Files Page - Teacher View in APTgt v2.0

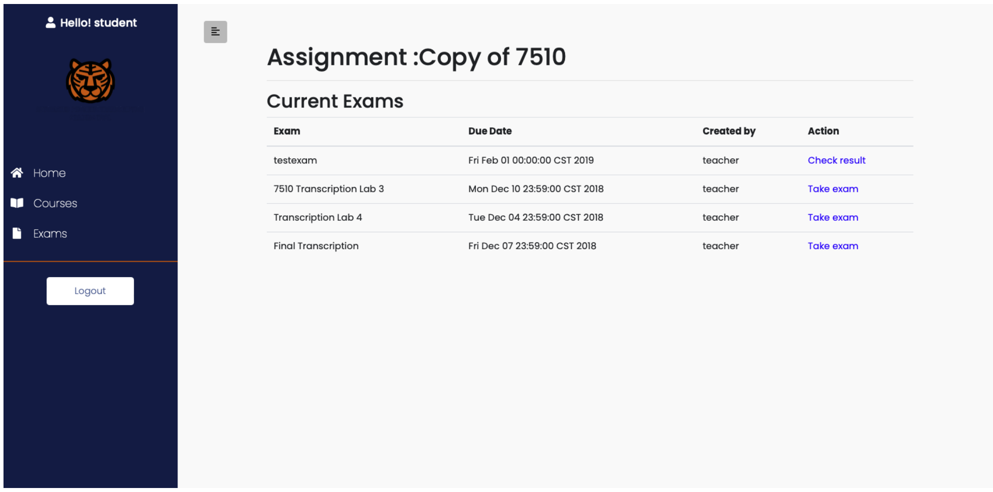


Figure 2.28: Exams Page - Student View in APTgt v2.0

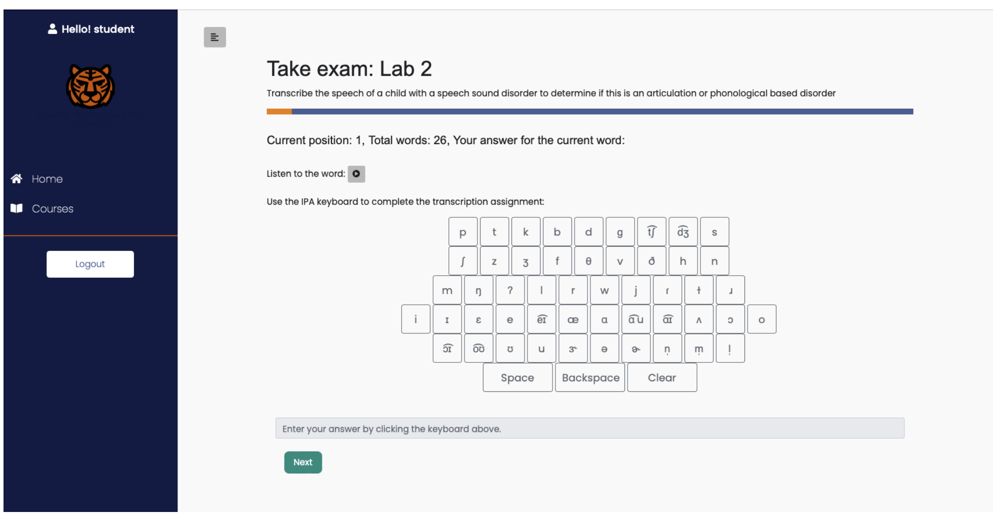
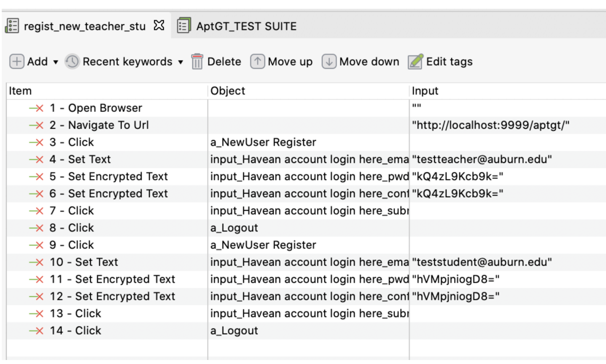


Figure 2.29: Take Exam - Student View in APTgt v2.0

cases and scripts. For example, figure 2.30 demonstrates the test cases to verify that a new user can successfully register as a teacher or a student following these test steps:

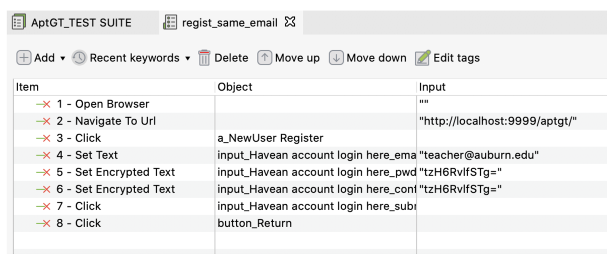
- Open Browser
- Open the application URL.
- Click on the "New User? Register" Link on the login page.
- Fill in the required registration information, such as username, and password.
- Select the role.
- Click on the "Submit" button.

The following figures illustrate our test suite, which consists of step-by-step test cases designed to evaluate the functionality of the system(see Figs. 2.31–2.43).



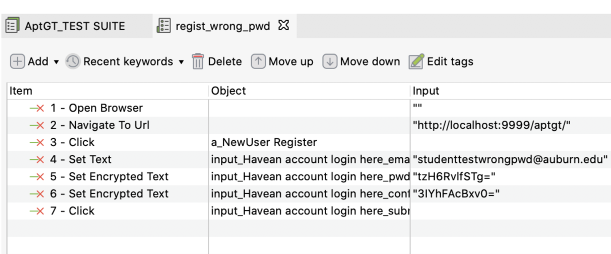
Item	Object	Input
1 - Open Browser		""
2 - Navigate To Url		"http://localhost:9999/aptgt/"
3 - Click	a_NewUser Register	
4 - Set Text	input_Havean account login here_ema	"testteacher@auburn.edu"
5 - Set Encrypted Text	input_Havean account login here_pwd	"kQ4zL9Kcb9k="
6 - Set Encrypted Text	input_Havean account login here_conf	"kQ4zL9Kcb9k="
7 - Click	input_Havean account login here_subı	
8 - Click	a_Logout	
9 - Click	a_NewUser Register	
10 - Set Text	input_Havean account login here_ema	"teststudent@auburn.edu"
11 - Set Encrypted Text	input_Havean account login here_pwd	"hVMpjniogD8="
12 - Set Encrypted Text	input_Havean account login here_conf	"hVMpjniogD8="
13 - Click	input_Havean account login here_subı	
14 - Click	a_Logout	

Figure 2.30: Test Case: New Registration



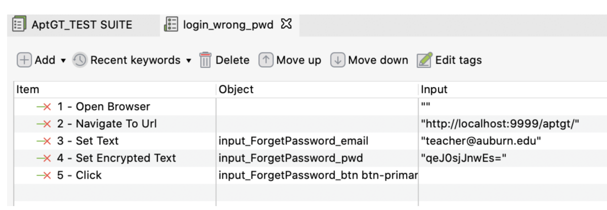
Item	Object	Input
1 - Open Browser		""
2 - Navigate To Url		"http://localhost:9999/aptgt/"
3 - Click	a_NewUser Register	
4 - Set Text	input_Havean account login here_ema	"teacher@auburn.edu"
5 - Set Encrypted Text	input_Havean account login here_pwd	"tzH6RvIfSTg="
6 - Set Encrypted Text	input_Havean account login here_conf	"tzH6RvIfSTg="
7 - Click	input_Havean account login here_subı	
8 - Click	button_Return	

Figure 2.31: Test Case: Repeated Registration



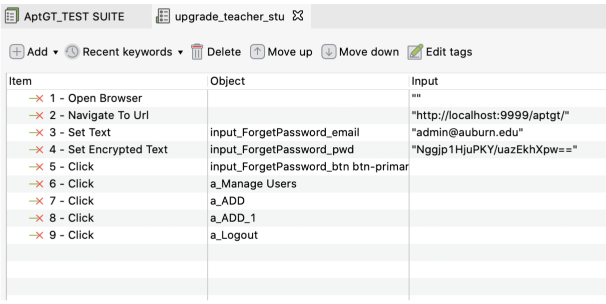
Item	Object	Input
1 - Open Browser		""
2 - Navigate To Url		"http://localhost:9999/aptgt/"
3 - Click	a_NewUser Register	
4 - Set Text	input_Havean account login here_ema	"studenttestwrongpwd@auburn.edu"
5 - Set Encrypted Text	input_Havean account login here_pwd	"tzH6RvIfSTg="
6 - Set Encrypted Text	input_Havean account login here_conf	"3IYhFAcBxv0="
7 - Click	input_Havean account login here_subı	

Figure 2.32: Test Case: Register with Wrong Password



Item	Object	Input
1 - Open Browser		""
2 - Navigate To Url		"http://localhost:9999/aptgt/"
3 - Set Text	input_ForgetPassword_email	"teacher@auburn.edu"
4 - Set Encrypted Text	input_ForgetPassword_pwd	"qeJ0sjJnwEs="
5 - Click	input_ForgetPassword_btn btn-primar	

Figure 2.33: Test Cases - Login with Wrong Password

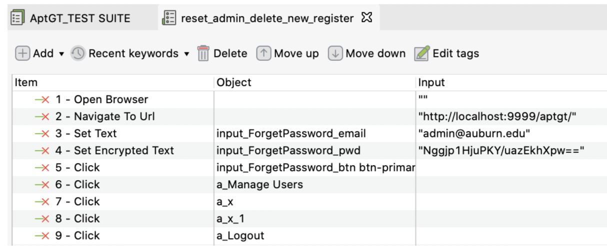


AptGT_TEST SUITE | upgrade_teacher_stu

Add | Recent keywords | Delete | Move up | Move down | Edit tags

Item	Object	Input
1 - Open Browser		""
2 - Navigate To Url		"http://localhost:9999/aptgt/"
3 - Set Text	input_ForgetPassword_email	"admin@auburn.edu"
4 - Set Encrypted Text	input_ForgetPassword_pwd	"Nggjp1HjuPKY/uazEkhXpw=="
5 - Click	input_ForgetPassword_btn btn-primar	
6 - Click	a_Manage Users	
7 - Click	a_ADD	
8 - Click	a_ADD_1	
9 - Click	a_Logout	

Figure 2.34: Test Case - Admin Add User

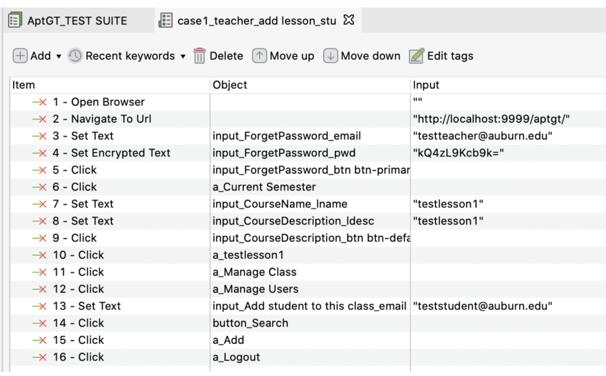


AptGT_TEST SUITE | reset_admin_delete_new_register

Add | Recent keywords | Delete | Move up | Move down | Edit tags

Item	Object	Input
1 - Open Browser		""
2 - Navigate To Url		"http://localhost:9999/aptgt/"
3 - Set Text	input_ForgetPassword_email	"admin@auburn.edu"
4 - Set Encrypted Text	input_ForgetPassword_pwd	"Nggjp1HjuPKY/uazEkhXpw=="
5 - Click	input_ForgetPassword_btn btn-primar	
6 - Click	a_Manage Users	
7 - Click	a_x	
8 - Click	a_x_1	
9 - Click	a_Logout	

Figure 2.35: Test Case: Admin Delete User

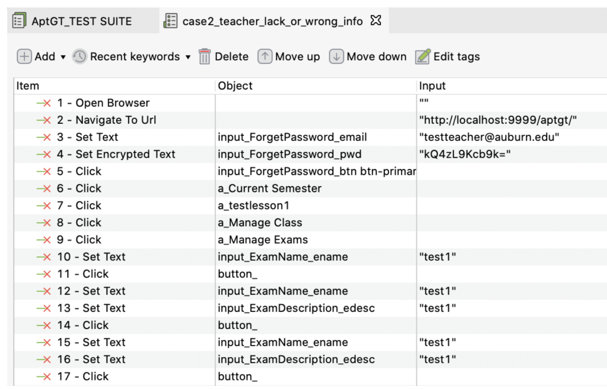


AptGT_TEST SUITE | case1_teacher_add lesson_stu

Add | Recent keywords | Delete | Move up | Move down | Edit tags

Item	Object	Input
1 - Open Browser		""
2 - Navigate To Url		"http://localhost:9999/aptgt/"
3 - Set Text	input_ForgetPassword_email	"testteacher@auburn.edu"
4 - Set Encrypted Text	input_ForgetPassword_pwd	"kQ4zL9Kcb9k="
5 - Click	input_ForgetPassword_btn btn-primar	
6 - Click	a_Current Semester	
7 - Set Text	input_CourseName_lname	"testlesson1"
8 - Set Text	input_CourseDescription_ldesc	"testlesson1"
9 - Click	input_CourseDescription_btn btn-defa	
10 - Click	a_testlesson1	
11 - Click	a_Manage Class	
12 - Click	a_Manage Users	
13 - Set Text	input_Add student to this class_email	"teststudent@auburn.edu"
14 - Click	button_Search	
15 - Click	a_Add	
16 - Click	a_Logout	

Figure 2.36: Test Case: Teacher Manage Class

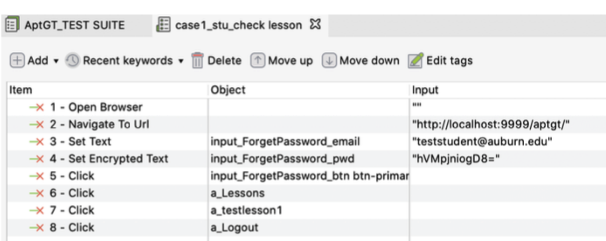


AptGT_TEST SUITE | case2_teacher_lack_or_wrong_info

Add | Recent keywords | Delete | Move up | Move down | Edit tags

Item	Object	Input
1 - Open Browser		""
2 - Navigate To Url		"http://localhost:9999/aptgt/"
3 - Set Text	input_ForgetPassword_email	"testteacher@auburn.edu"
4 - Set Encrypted Text	input_ForgetPassword_pwd	"kQ4zL9Kcb9k="
5 - Click	input_ForgetPassword_btn btn-primar	
6 - Click	a_Current Semester	
7 - Click	a_testlesson1	
8 - Click	a_Manage Class	
9 - Click	a_Manage Exams	
10 - Set Text	input_ExamName_ename	"test1"
11 - Click	button_	
12 - Set Text	input_ExamName_ename	"test1"
13 - Set Text	input_ExamDescription_edesc	"test1"
14 - Click	button_	
15 - Set Text	input_ExamName_ename	"test1"
16 - Set Text	input_ExamDescription_edesc	"test1"
17 - Click	button_	

Figure 2.37: Test Case: Teacher Manage Exam

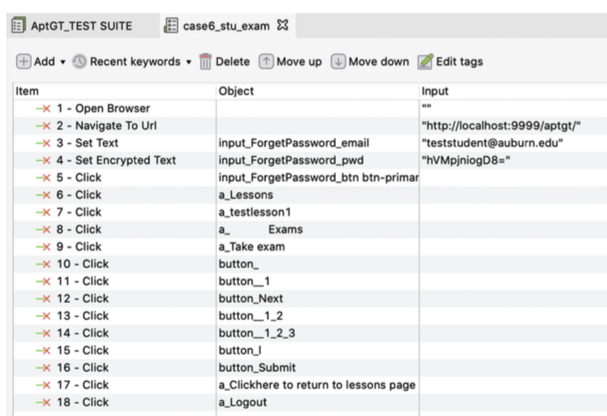


AptGT_TEST SUITE | case1_stu_check lesson

Add | Recent keywords | Delete | Move up | Move down | Edit tags

Item	Object	Input
1 - Open Browser		""
2 - Navigate To Url		"http://localhost:9999/aptgt/"
3 - Set Text	input_ForgetPassword_email	"teststudent@auburn.edu"
4 - Set Encrypted Text	input_ForgetPassword_pwd	"hVMpjniogD8="
5 - Click	input_ForgetPassword_btn btn-primar	
6 - Click	a_Lessons	
7 - Click	a_testlesson1	
8 - Click	a_Logout	

Figure 2.38: Test Case: Student Check Course



AptGT_TEST SUITE | case6_stu_exam

Add | Recent keywords | Delete | Move up | Move down | Edit tags

Item	Object	Input
1 - Open Browser		""
2 - Navigate To Url		"http://localhost:9999/aptgt/"
3 - Set Text	input_ForgetPassword_email	"teststudent@auburn.edu"
4 - Set Encrypted Text	input_ForgetPassword_pwd	"hVMpjniogD8="
5 - Click	input_ForgetPassword_btn btn-primar	
6 - Click	a_Lessons	
7 - Click	a_testlesson1	
8 - Click	a_ Exams	
9 - Click	a_Take exam	
10 - Click	button_	
11 - Click	button__1	
12 - Click	button_Next	
13 - Click	button__1_2	
14 - Click	button__1_2_3	
15 - Click	button_I	
16 - Click	button_Submit	
17 - Click	a_Clickhere to return to lessons page	
18 - Click	a_Logout	

Figure 2.39: Test Case: Student Take Exam

Figure 2.40: Test Case: Teacher Use IPA Keyboard

Figure 2.41: Test Case: Student Use IPA Keyboard

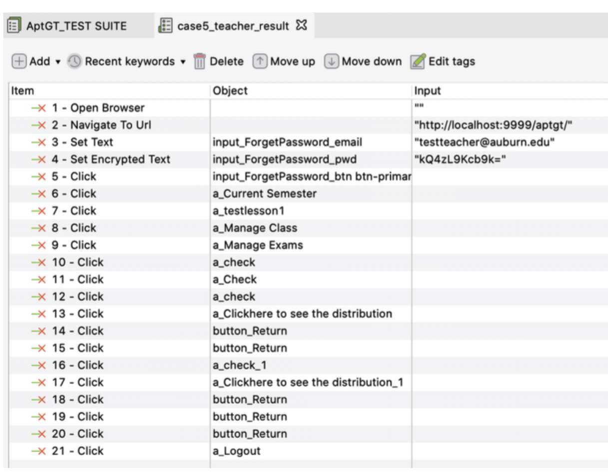


Figure 2.42: Test Case: Teacher Check Exam Result

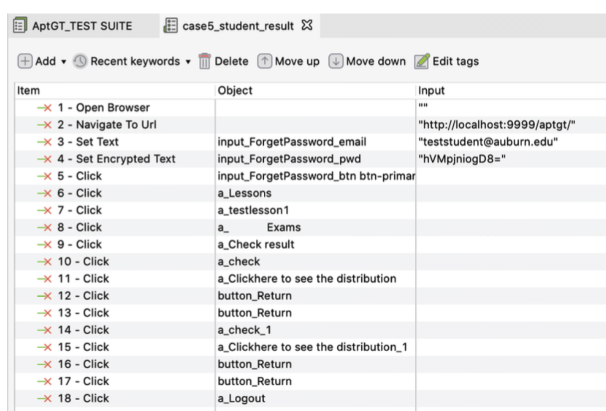


Figure 2.43: Test Case: Student Check Exam Result

As a collaborative team, we leveraged the cloud dashboard in Katalon Studio to facilitate our testing project. We created a total of thirty-nine test cases to thoroughly test the newly designed system, with each case covering multiple roles: admin, teacher, and student. The dashboard provided a centralized location to track the progress of our test cases, view detailed test results, and identify any issues or failures. The dashboard below displays the test results(see Fig. 2.44), indicating that all the test cases have been executed and passed successfully.

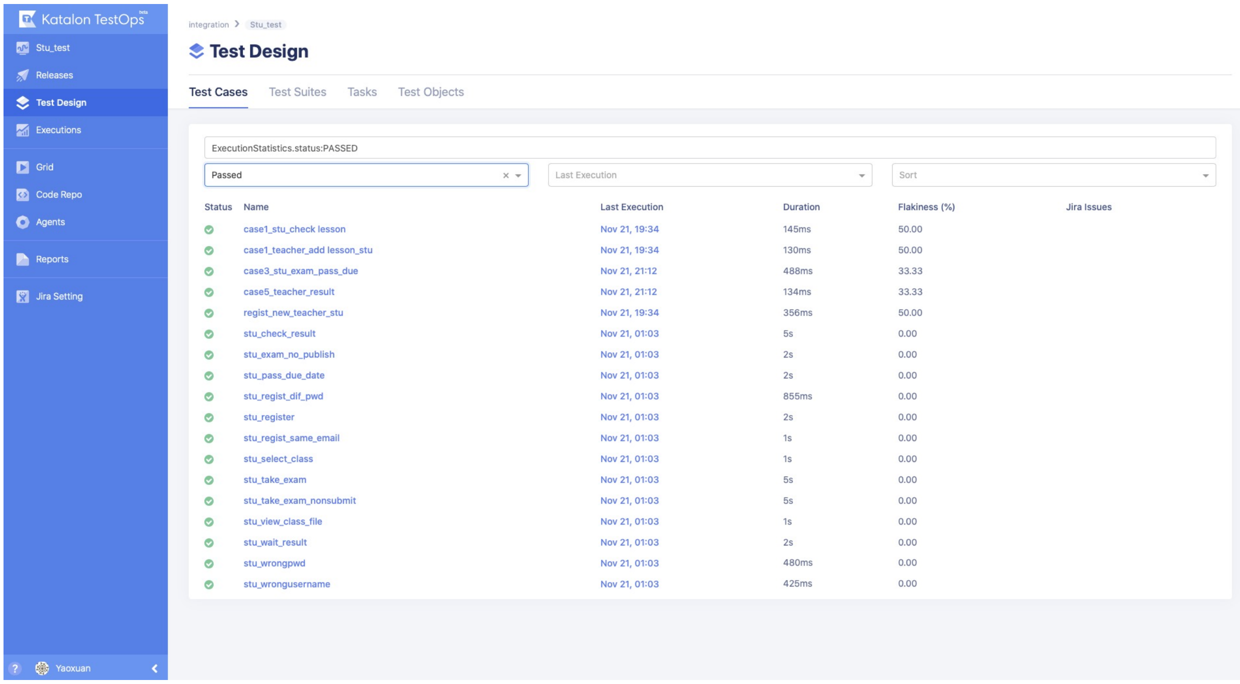


Figure 2.44: Test Cases in Katalon Dashboard

2.5 Evaluation & Result

2.5.1 Participants

For this study, a small group of linguistic professionals and usability experts were selected to conduct comparative usability testing. For linguistic professional selection, our minimum criterion is advanced undergraduate, graduate, or faculty in a field that is practiced in linguistic transcription. For usability expert selection, our minimum criterion will be usability certification, a course in usability, or a usability expert that is a current usability researcher or member of technical staff.

2.5.2 Procedure & Data Collection

Study 1 was conducted with Institutional Review Board (IRB) approval, and an information letter was presented to participants prior to their participation in the study. The approved IRB documentation can be found in Appendix A. We used online questionnaires in Qualtrics to gather data from participants. First, the contextual information and eligibility data were collected from study participants on their background and experience in Linguistics and E-learning in the pre-questionnaire. The participants were asked about their general information such as gender, age, professional experience, and experience in E-learning and usability test, etc. After the pre-survey, the participants were provided the system with the study details, the access link of two versions of the application, and the task the participant will have to complete. Tasks for the experimental interface and benchmark interface were the same. Participants were asked to perform various tasks such as logging in as a teacher and student, exploring different sections of the system, accessing specific exams, and viewing exam results in both versions. The post-questionnaire was created to collect the participant's feedback on the system in the aspects of aesthetics, ease of use, satisfaction, and overall usability to verify our hypotheses.

2.5.3 Result & Analysis

Pre-Survey

In the pre-survey, a total of 21 participants, consisting of undergraduate and graduate students, as well as professionals from Auburn University, were involved. Among the participants, 14 were female and seven were male. Two participants were undergraduate students, 13 were graduate students, and six were professionals either in linguistics or usability. It was found that 76% of the participants had prior experience working with E-learning systems, with the most commonly used platforms being Canvas, Coursera, and Udemy.

One of the survey questions, Q9, asked participants to identify any specific challenges they encountered when using E-learning systems. The results, shown in Figure 2.45, revealed that the most commonly selected challenge was “Unappealing”, with a percentage of 24.39%. Additionally, 17.07% of participants reported challenges related to “Inconsistency design” and ”Navigation difficulties.” Furthermore, 12.2% of participants highlighted issues related to “Poor use of colors/typography/layout”.

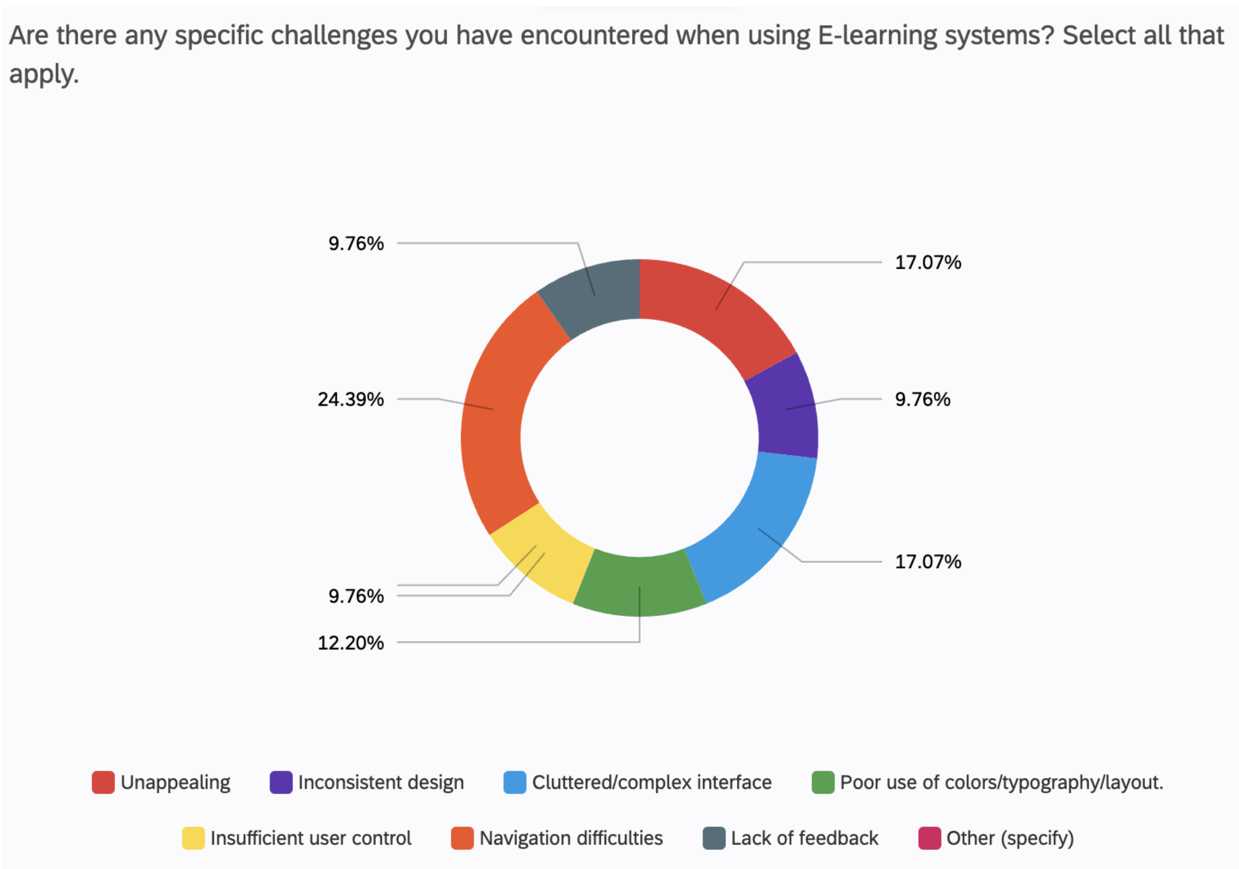


Figure 2.45: Pre-survey Result - Challenges Encountered When Using E-learning Systems

Among the participants with prior E-learning system experience, five out of 16 had used the APTgt system (APTgt v1.0) before. They were asked a specific question, Q13, regarding challenges encountered when using APTgt. The result, depicted in Figure 2.46, indicated that 21.43% of these participants found the old APTgt design to be unappealing, inconsistent in design, and difficult to navigate.

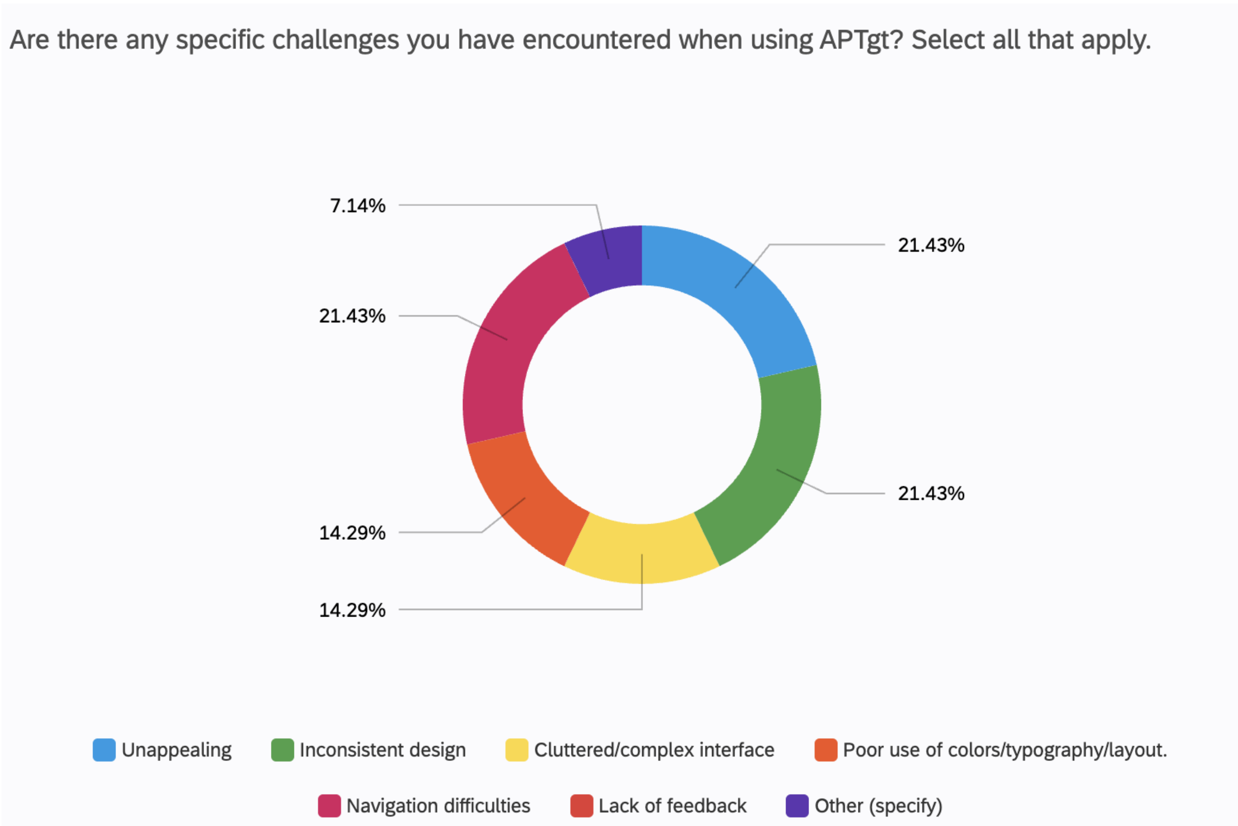


Figure 2.46: Pre-survey Result - Challenges Encountered When Using APTgt System (APTgt v1.0)

Regarding prior usability testing experience, 13 out of the 21 participants reported having such experience, with the most common method being survey/questionnaire-based evaluations.

The above statistics provide valuable insight into the general challenges encountered by users of E-learning systems and specifically highlight the issues associated with the user interface of APTgt v1.0. These findings further reinforce the motivation behind our research to optimize the system's user interface design in terms of aesthetics, consistency, and overall satisfaction. Moreover, the chosen usability evaluation method aligns well with the participants' prior experience in this area.

Post-Survey

After the completion of the pre-survey session, 12 out of 21 participants demonstrated their interest in participating in the follow-up tasks and post-survey. Among these participants, there were six usability experts, five linguistic professionals, and one individual specializing in software engineering, who had also taken a user interface design course. Subsequently, the tasks and instructions, along with the post-survey link, were distributed via the provided email addresses of the participants.

The post-survey results of APTgt v2.0 revealed positive feedback from the participants, indicating a favorable experience with the updated version compared to the old version. When asked about their overall experience with APTgt v2.0 in comparison to APTgt v1.0, the majority of participants expressed a positive sentiment. The responses were collected on a scale of 1-5, with 'Strongly Disagree' representing the lowest rating and 'Strongly Agree' representing the highest rating. The average rating for the questions pertaining to the overall reaction to APTgt v2.0 was 4.54, as shown in Figure 2.47.

#	Question	Strongly Agree		Agree		Neutral		Disagree		Strongly Disagree	Weighted Average	Total
1	The application is more visually pleasing than old version.	75.00%	9	16.67%	2	0.00%	0	8.33%	1	0.00%	**4.58**	12
2	The application is easier to use than old version.	58.33%	7	33.33%	4	8.33%	1	0.00%	0	0.00%	**4.5**	12
3	The navigation is more efficient than old version.	66.67%	8	25.00%	3	8.33%	1	0.00%	0	0.00%	**4.58**	12
4	I am more satisfied with the application than the old version.	58.33%	7	33.33%	4	8.33%	1	0.00%	0	0.00%	**4.5**	12
											4.54	

Figure 2.47: Post-survey Result - Ratings of APTgt v2.0 compared to the old version

When participants were asked to rate the usability of APTgt v2.0 in terms of layout, use of color, consistency, ease of use, and navigation, the average rating received was 4.48 on a scale of 1-5, where 'Very Poor' represented the lowest rating and 'Excellent' represented the highest rating. This indicates a positive evaluation of the mentioned usability aspects in APTgt v2.0. The corresponding table presents the results visually as shown in Figure 2.48.

The null hypotheses for our study stated that there would be no significant difference between the experimental version (APTgt v2.0) and the benchmark version (APTgt v1.0) in terms of "Aesthetically pleasing", "Ease of use", and "Satisfaction"(refer to Section 2.1.4).

#	Question	Excellent		Very Good		Good		Poor		Very poor	Weighted Average	Total
1	Layout	58.33%	7	8.33%	1	33.33%	4	0.00%	0	0.00%	**4.25**	12
2	Use of color	25.00%	3	50.00%	6	25.00%	3	0.00%	0	0.00%	**4**	12
3	Consistency	83.33%	10	16.67%	2	0.00%	0	0.00%	0	0.00%	**4.83**	12
4	Easy to use	50.00%	6	50.00%	6	0.00%	0	0.00%	0	0.00%	**4.5**	12
5	Navigation	83.33%	10	16.67%	2	0.00%	0	0.00%	0	0.00%	**4.83**	12
											4.48	

Figure 2.48: Post-survey Result - Usability rating of APTgt v2.0 in terms of various aspects

However, the survey results from our participants provided compelling evidence to reject these null hypotheses. The majority of participants consistently expressed a higher level of usability in APTgt v2.0 across multiple aspects. The improvements in the new version were particularly evident in terms of aesthetics (layout, use of color, and consistency), ease of use (efficient navigation and absence of difficulties), and higher user satisfaction (likelihood of reusing the application and recommending it to others).

Specifically, a majority of participants, 91.67%, agreed that the experimental system exhibited a higher level of visual appeal compared to the benchmark system, as indicated in Figure 2.49. Moreover, all participants provided above-average ratings for aspects related to layout, use of color, and consistency, further confirming the aesthetic appeal of the new version system, as shown in Figure 2.50. These results convincingly rejected the null hypothesis in the "Aesthetically pleasing" aspect, indicating a clear difference between the experimental and benchmark interfaces.

In terms of ease of use, the majority of participants (91.66%) agreed that APTgt v2.0 is easier to use compared to the old version, as indicated in Figure 2.47. Additionally, all participants agreed that the new version helped them complete tasks more efficiently, with 75% strongly agreeing(see Fig. 2.51). Furthermore, 91.67% of individuals reported no difficulties while using APTgt v2.0(see Fig. 2.52). These findings strongly refute the null hypothesis regarding "Ease of use".

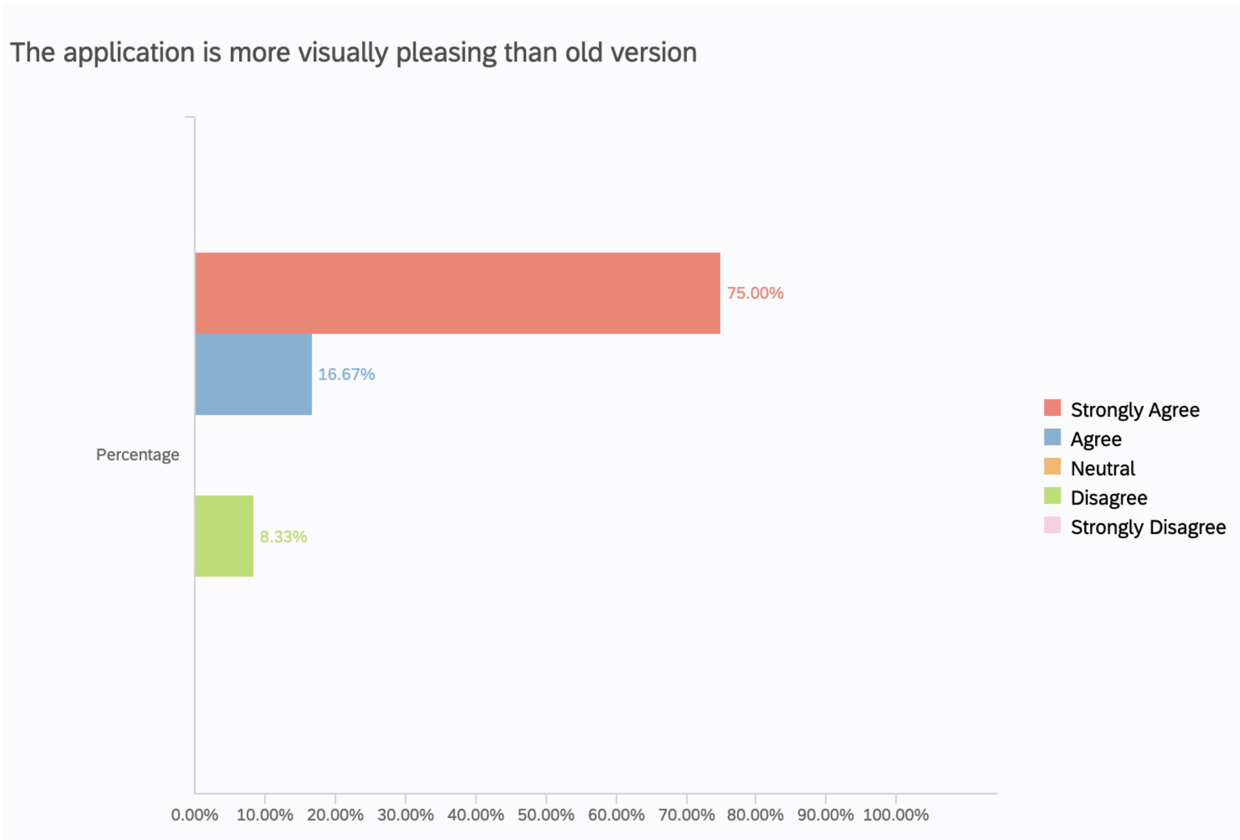


Figure 2.49: Post-survey Aesthetics Result

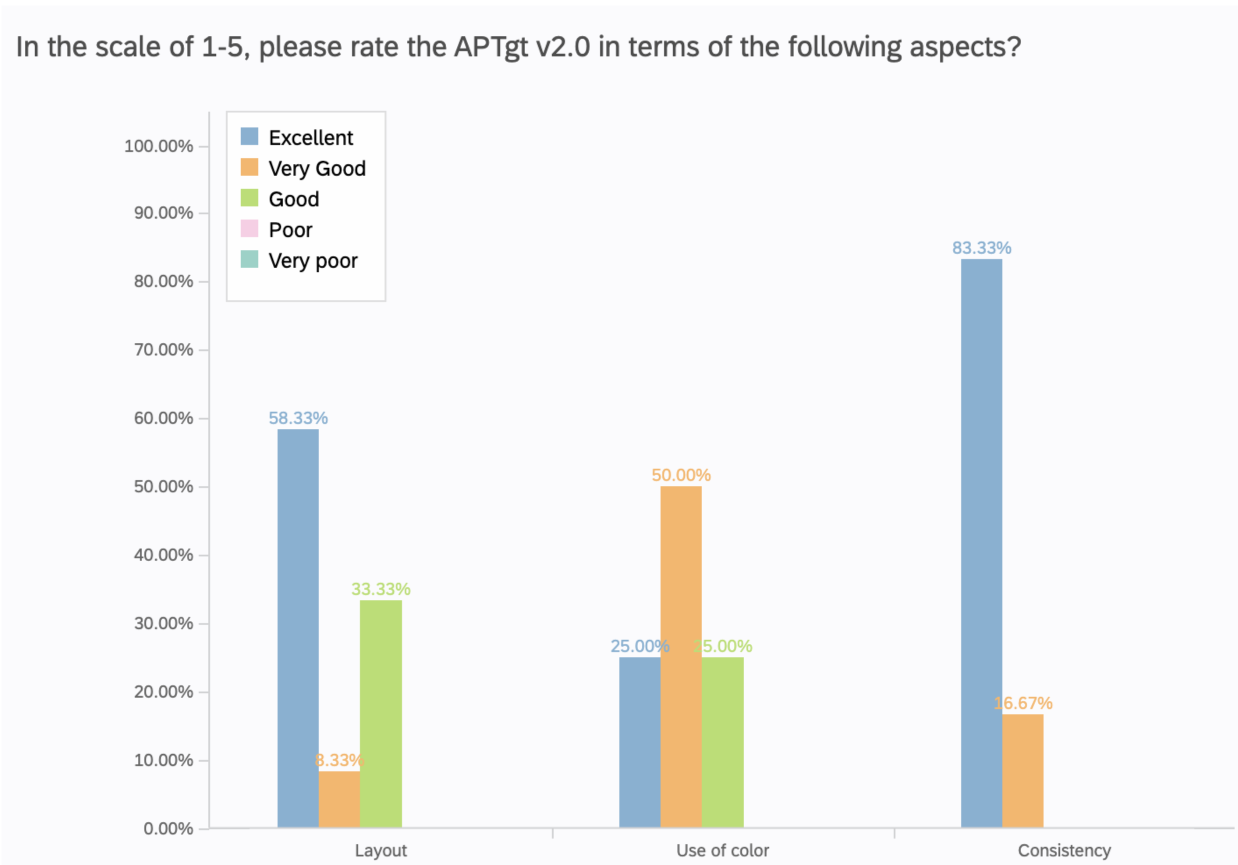


Figure 2.50: Post-survey Result - Usability rating of APTgt v2.0 in terms of layout, use of color and consistency

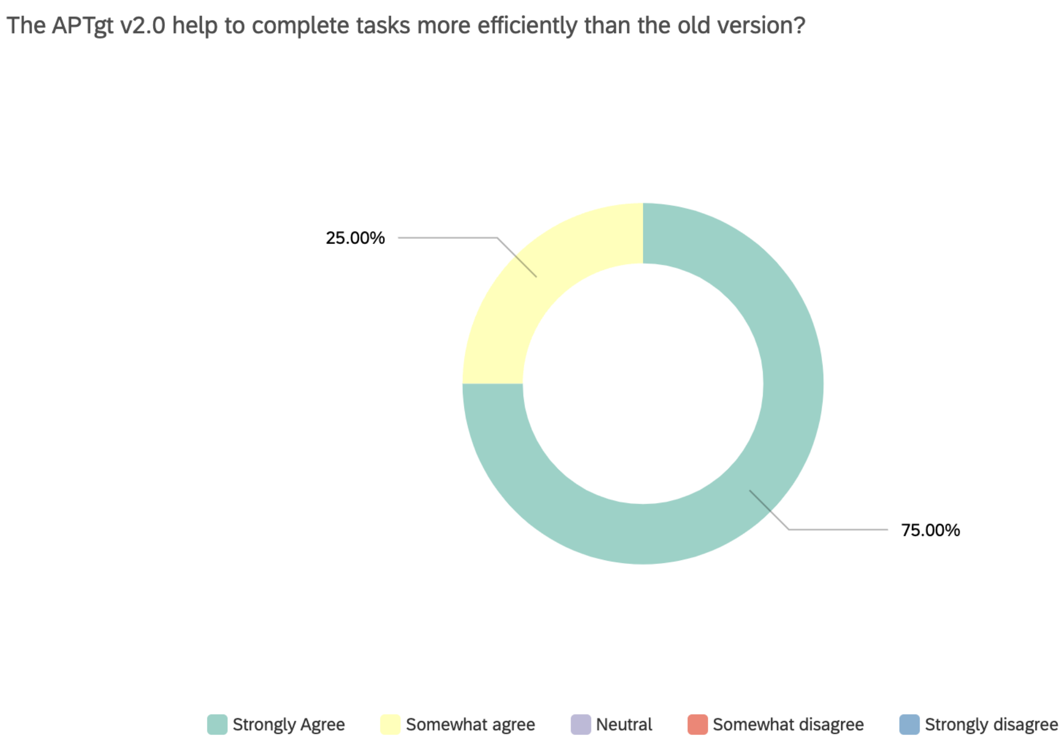


Figure 2.51: Post-survey Efficiency Result

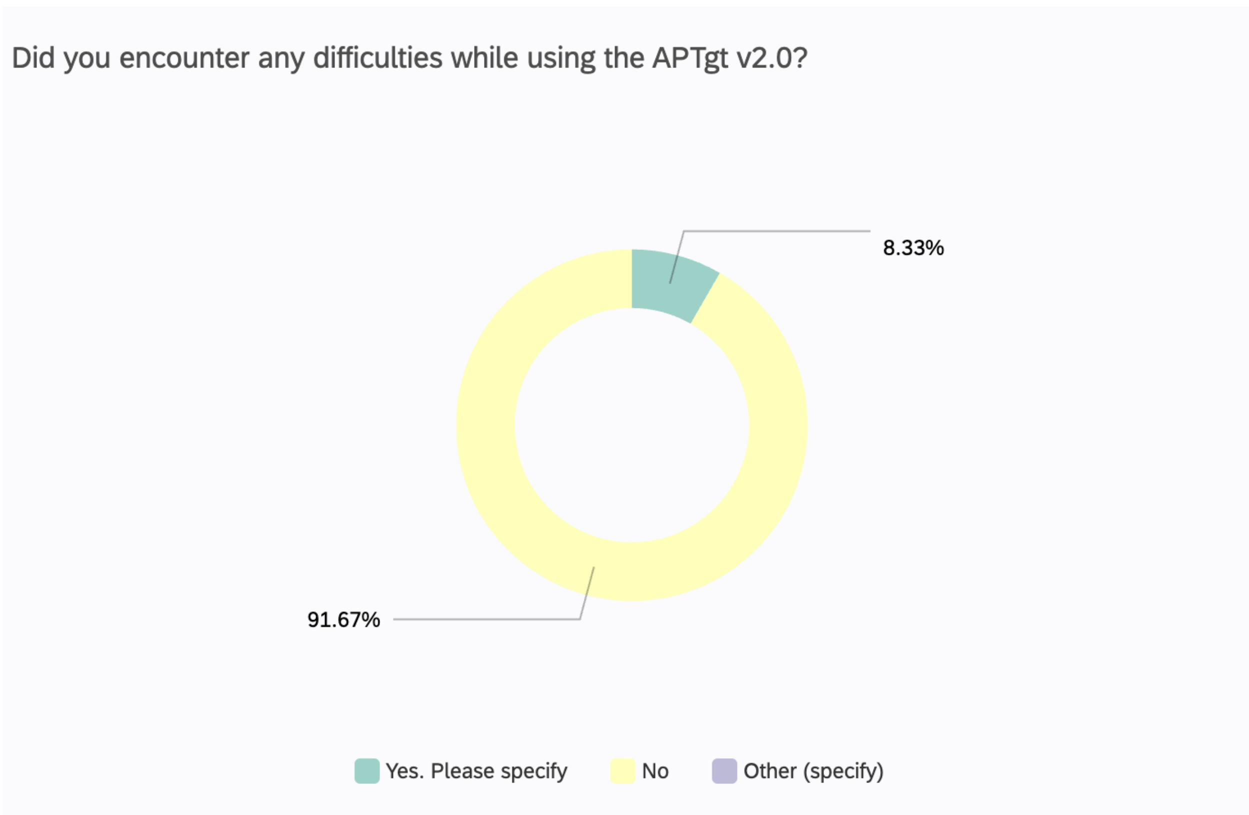


Figure 2.52: Post-survey Difficulties Encountered Result

The results regarding satisfaction also explicitly rejected the null hypothesis, as participants showed a high likelihood of recommending APTgt v2.0 to others. The average rating of **8.59** out of 10, as calculated from the data points depicted in Figure 2.53, further supports this positive sentiment. In addition, no participants reported the absence of any aspects or features from APTgt v1.0 in APTgt v2.0(see Fig. 2.54), indicating a successful transition and inclusion of desired elements. For the likelihood of continued use, 61% of participants expressed their intention to continue using APTgt v2.0 in their teaching or professional activities(see Fig. 2.55). Particularly, figure 2.56 shows that linguistic professionals, our target user audience, expressed definite interest in using the application (definitely or probably will). A rating of 4.2 out of 5 on the likelihood of continued use from target clients indicates a high level of satisfaction with the application.

Figure 2.53: Post-survey Result - Likelihood of Recommending to Others

2.6 Conclusion

In conclusion, this study focused on optimizing the user interface and user experience of the Automated Phonetic Transcription Grading Tool to provide an improved and efficient

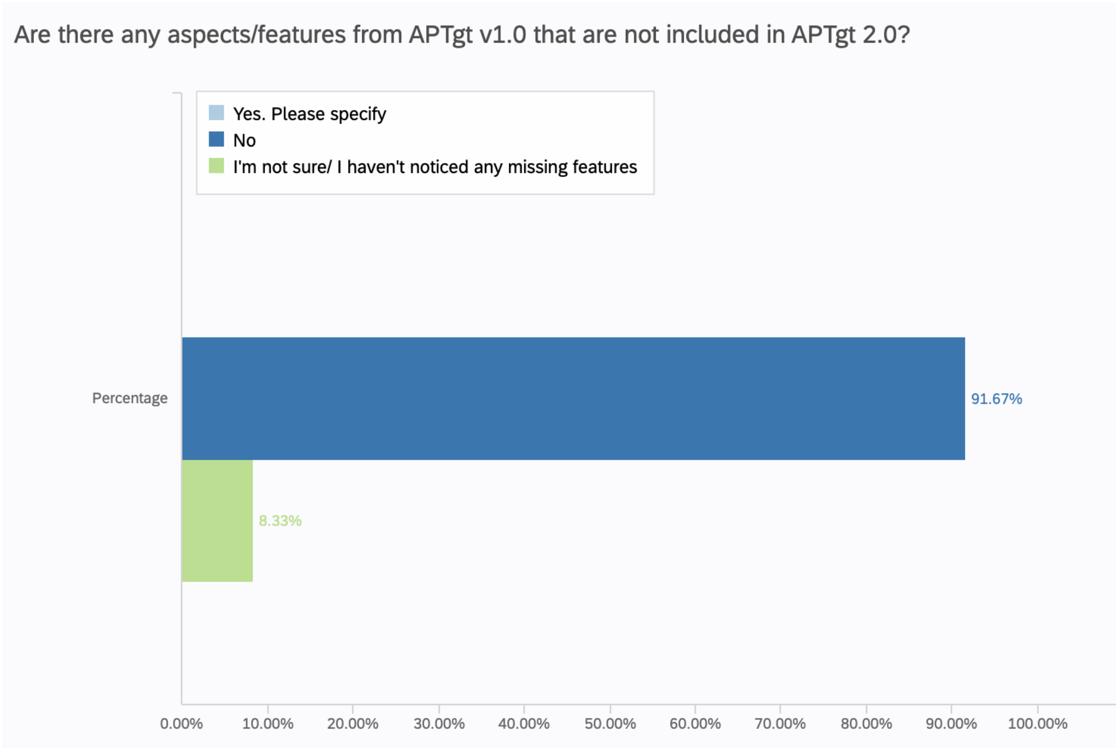


Figure 2.54: Post-survey Absence of Features Result

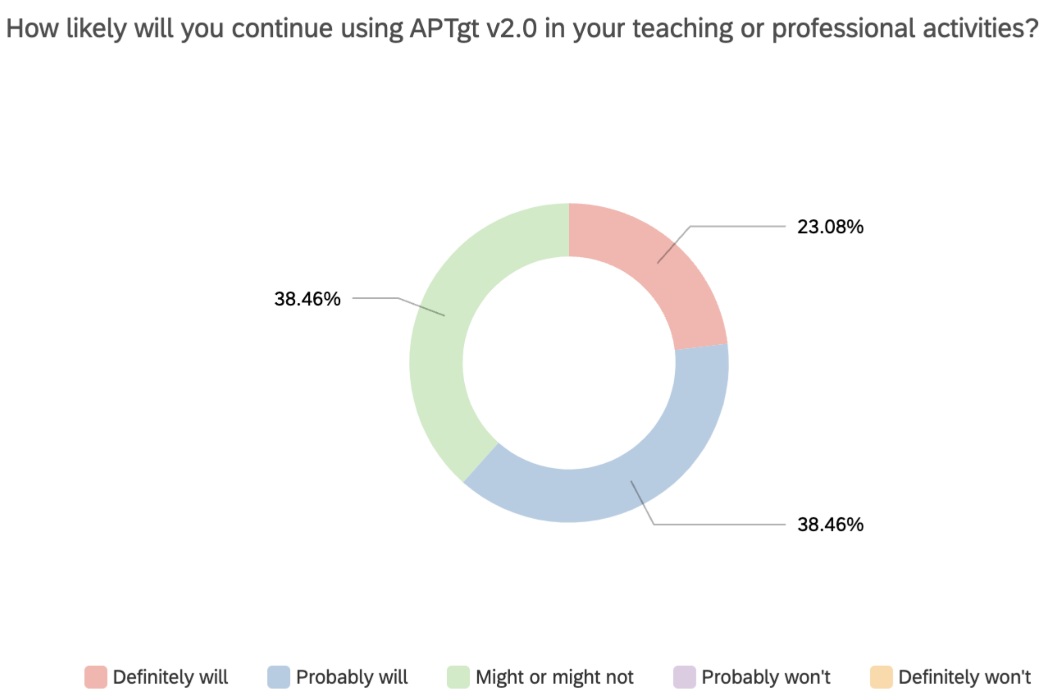


Figure 2.55: Post-survey Result - Likelihood of Continued Use

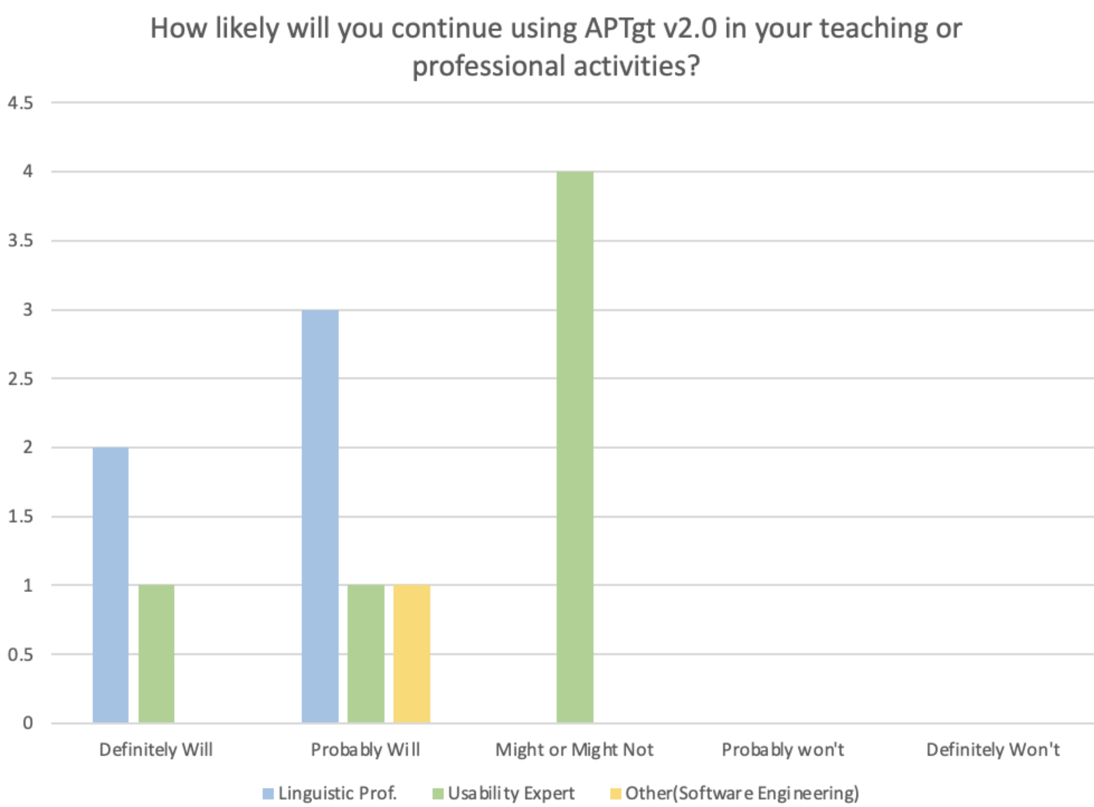


Figure 2.56: Post-survey Result - Likelihood of Continued Use Based on Role

tool for communication disorders faculty. The research addressed the design and usability concerns identified in the previous iterations of the platform, aiming to enhance ease of use, aesthetics, consistency in design, and overall usability. The post-survey results demonstrated positive feedback from participants, indicating that APTgt v2.0 had a higher level of usability, and user satisfaction compared to the previous version. These improvements in the user interface and user experience of APTgt contribute to the field of phonetic E-learning by providing linguistic instructors with a more effective and user-friendly tool for phonetic transcription training and exam generation.

The APTgt system has undergone a significant update to Aptgt 2.0, which has now been deployed on the online server. The development process has been informed by valuable feedback from Linguistic instructors and students, allowing for iterative design improvements. As new requirements arise, the system will continue to be updated and expanded, ensuring that new features and functionality align seamlessly with the current version.

Chapter 3

Transformer-based Multilingual G2P conversion

3.1 Introduction

3.1.1 Background

Phonetic transcription is representative of speech sounds in specific symbols, primarily through the International Phonetic Alphabet (IPA)[7, 4]. In the field of communication disorders, the clinical phonetic transcription skill is a critical part of students' clinical preparation to become speech-language pathologists. However, students often report feeling unprepared to apply the skill in clinical practice as the practice opportunities can be impeded by the limited phonetic learning resources. In recent years, technological advancements have attempted to address this issue by leveraging computers and digital tools to bring diverse speech samples into the classroom and provide additional transcription practice[48].

The APTgt system aims to support phonetic learning by offering interactive IPA-based phonetic transcription exams, utilizing machine learning technology to automate and optimize the system. One of the key features of our system is the auto exam generator (see Fig. 3.1), designed to assist linguistic instructors in effortlessly creating phonetic exams [35]. With just a single click, the system retrieves audio questions from a pre-stored word

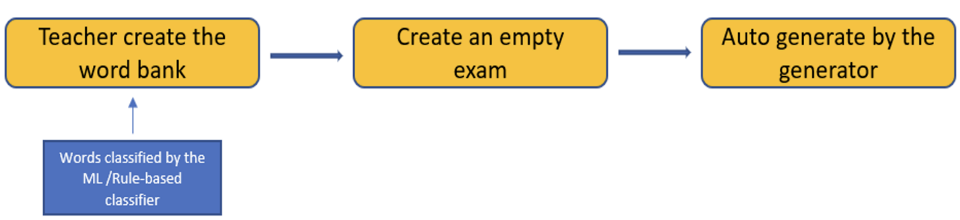


Figure 3.1: The workflow of APTgt Auto Exam Generator

bank[33](See Fig. 3.2) and generates a phonetic exam with appropriate difficulty levels.

These difficulty levels are determined by a multi-class classification module, ensuring that the exams align with the student's proficiency levels.

Word Bank

Word File		File Pronunciation	Difficulty Score	Difficulty Level	Listen	Edit	
1597114369971_file_example_WAV_1MG.wav	θvrɚf̪uwr	θvrɚf̪uwr	40.0	adv 1	Listen	Edit	x
1598327296330_file_example_WAV_1MG.wav	u˥˧	u˥˧	18.5	medium 1	Listen	Edit	x
1598327856027_file_example_WAV_1MG.wav	eɪ͡eeɪ	eɪ͡eeɪ	17.0	medium 1	Listen	Edit	x
1598327869819_file_example_WAV_1MG.wav	bʃpv̋	bʃpv̋	9.5	easy 2	Listen	Edit	x
1598327883683_file_example_WAV_1MG.wav	ml̩ṳ˥	ml̩ṳ˥	20.5	medium 1	Listen	Edit	x
1598327929896_file_example_WAV_1MG.wav	ənɽho̠˥	ənɽho̠˥	38.5	adv 1	Listen	Edit	x
1598327940111_file_example_WAV_1MG.wav	ʌaɪ͡leɪ͡e	ʌaɪ͡leɪ͡e	19.5	medium 1	Listen	Edit	x
1598327956548_file_example_WAV_1MG.wav	hvʒɔaʊ͡ɛn̩	hvʒɔaʊ͡ɛn̩	14.5	medium 1	Listen	Edit	x
1598327971578_file_example_WAV_1MG.wav	ɾθʔæʌm̩ʊ̃	ɾθʔæʌm̩ʊ̃	26.5	medium 2	Listen	Edit	x
1598327981255_file_example_WAV_1MG.wav	ðbz	ðbz	3.0	easy 1	Listen	Edit	x

Figure 3.2: APTgt Word Bank

3.1.2 Research Problem & Motivation

The current APTgt system can classify disordered and non-disordered speech and identify question difficulty levels automatically. However, it relies on a pre-stored word bank to function, which instructors must manually create by generating the word or phrase pronunciation as the corresponding correct answers for audio questions in advance. Unfortunately, this requires instructors to manually convert words to IPA format, which can be inefficient and prone to typing errors. Compared with manual transcription, auto-generating can be much more effective because it can handle large amounts of data quickly and accurately, without the need for human input.

To support the efficient phonetic transcription process in the phonetic exams generation of our linguistic E-learning system APTgt, we designed a neural machine translation (NMT)[53, 55] tool to translate English words (grapheme sequence) to their corresponding IPA phonetic spelling formats (phoneme sequence). For example, the word "*brown*" should be converted to "/braʊn/". This can be done by utilizing the Grapheme-to-Phoneme (G2P) technique[13, 14]. It will help instructors bring diverse speech samples into the classroom, and to provide more transcription practice opportunities for students.

3.1.3 Research Questions

The following research questions will guide this study:

1. What kind of machine learning technologies can be employed?
2. What performance it can achieve?
3. What methods can be adopted to improve the accuracy of translations?

3.1.4 Research Hypothesis

1. The G2P converter is expected to significantly improve the efficiency of generating phonetic transcriptions and facilitate the process of exam creation. This is based on the assumption that it can accurately convert graphemes to phonemes, which will reduce the manual effort required for transcription.

   H1: There will be a significant increase in APTgt efficiency utilizing the G2P converter.

2. To make our E-learning system more functional and expand it to support multiple languages, the converter is also expected to accurately convert text to phonemes for different languages.

   H2: There will be more languages supported with this version of the system than the prior APTgt system.

In this study, we selected the Transformer[51], a prominent deep learning model that has been widely adopted in natural language processing (NLP), computer vision (CV), and speech processing[32], to build our G2P neural machine translator. Also, to improve the functionality of the E-learning system and prepare for the expansion of this system to a multilingual system, we trained multiple language models and generated a multilingual G2P translator. Moreover, we evaluated our G2P system by the metrics of word error rate (WER) and phoneme error rate (PER)[10, 13].

The main contribution of the proposed G2P converter is to enhance the phonetic exam E-learning system and help linguistic instructors more comfortably and efficiently generate phonetic transcriptions by eliminating the majority of manually input. With this effort, the word bank can grow from one hundred to thousands of words and more using an automated process. Additionally, the G2P converter can be used to annotate the speech samples in the new or existing speech corpus for phone-level automatic speech recognition tasks.

3.2 Literature Review

Dr. Rao's work on Grapheme-to-phoneme (G2P) conversion using Long Short-Term Memory (LSTM) recurrent neural networks is relevant to our research problem[40]. In linguistics, a grapheme is the smallest unit of a written language, while a phoneme is the smallest unit of speech sound. A grapheme-to-phoneme system converts a spelled-out word to its phonetic format, making it useful in applications such as text-to-speech systems, where natural-sounding speech requires correct pronunciation of words. In his paper, Dr. Rao proposed a G2P model based on LSTM RNNs. The model was evaluated through several experiments, including unidirectional LSTM (ULSTM) with different output delays and deep bidirectional LSTM (DBLSTM) with a connectionist temporal classification (CTC) layer. The CTC output layer interprets the network outputs as a probability distribution over all possible output label sequences, conditioned on the input data. The CTC objective function directly maximizes the probabilities of the correct labelings. Dr. Rao used the public CMU pronunciation dictionary to train the G2P model, and the word error rate (WER) was used to evaluate its performance. The best reported (to our knowledge) WER on the public CMU dataset was achieved by combining the DBLSTM-CTC model with a traditional n-gram approach, resulting in a WER of 21.3%.

In the paper "A survey of deep learning techniques for neural machine translation"[55], Yang first provides an overview of NMT and its history and describes the development of NMT from traditional rule-based and statistical machine translation (SMT) to modern NMT models. Yang also discusses the different neural network architectures used in NMT, including recurrent neural networks (RNNs), convolutional neural networks (CNNs), and transformer models. CNN-based models have advantages in training speed and resolving the gradient vanishing problem. However, they have two fatal drawbacks that affect their translation quality: 1) they can only capture word dependencies within the width of their filters, which often leads to worse performance than RNN-based models for long dependencies; and 2) they compress sentences into a fixed-size vector, leading to a large performance reduction for longer sentences due to limited representation ability. In contrast, RNNs and transformers can capture long-term dependencies in the input sequence. RNN-based models were the first to achieve good results in NMT but have been largely surpassed by Transformer-based models. Transformer-based models rely on an enhanced version of the Attention Mechanism, called Self-Attention, to achieve state-of-the-art performance state-of-the-art results on many NMT benchmarks, especially on long sentences. The Self-Attention mechanism allows the model to read the entire sentence and model it at once, making it a powerful feature extractor with high inference speed. This feature makes the Transformer a combination of the advantages of both CNN and RNN models, giving it good feature representation ability. The innovative attention structure is the key to the Transformer's significant improvement in performance. Overall, the paper suggests that Transformer-based models are currently the most effective for NMT, but that there is ongoing research to explore other architectures and improve the performance of existing models.

In the paper "Neural Machine Translation for Multilingual Grapheme-to-Phoneme Conversion"[46], the model architecture is also an LSTM-based G2P model but evaluated on multilingual datasets including English, French, German, Dutch, and Spanish. The performance of the model was assessed using word error rate and phoneme error rate metrics at both the word and phoneme levels. The results are summarized in table 3.1. We consider this model as the baseline for comparing the performance of our proposed method, using the same metrics.

Table 3.1: Performance Comparison between Baseline and our Transformer-based Multilingual Model

	LSTM G2P[46]	
Language	PER(%)	WER(%)
English	12.9	52.8
French	7.7	43.1
Spanish	5.3	36.4
Dutch	2.8	13.5
German	4.2	18.4

Vaswani et al. establish a new model architecture of Transformers[51], employing an attention mechanism instead of recurrence, leading to state-of-the-art translation quality. By using multiple attention distributions and multiple outputs for a single input, the Transformer improves its performance on various tasks. Additionally, the use of layer normalization and residual connections simplifies optimization, enhancing the overall efficiency and effectiveness of the model.

The structure of a transformer (See Figure 3.3) consists of an encoder on the left and a decoder on the right. The encoder includes two sub-layers, namely the Multi-Head Attention and the Feed Forward network, with each sub-layer connected by a residual connection followed by layer normalization. The encoder is responsible for embedding the source sentence with word vectors and performing encoding. To maintain the sequence order, positional encoding is employed to determine the relative/absolute positions of each token in the sequence. In the encoder, self-attention layers are employed, where the keys, values, and queries all originate from the same source. It allows the model to attend to all previous layers of encoding.

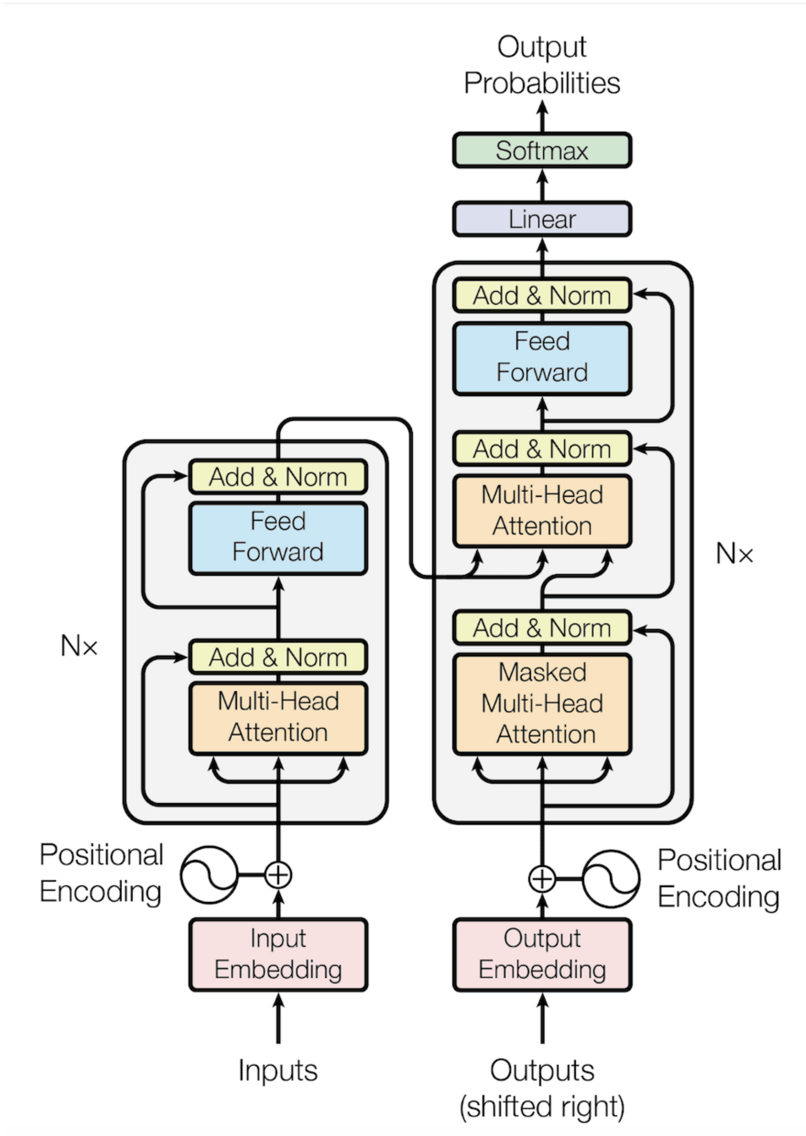


Figure 3.3: Transformer Model Architecture

The decoder model in the Transformer architecture is similar to the encoder, but it includes a Masked Multi-Head Attention mechanism. This attention mechanism attends to the previous decoder states and masks the future tokens during word decoding to prevent repetition in target sentences. This ensures that each word is generated based only on the previously decoded words and not on future words. The Attention mechanism of the Transformer also used a Scaled Dot-Production Attention approach. This involves computing the dot product of keys and queries, divided by the square root of the keys' dimension, to obtain the weights of the values. The softmax function is then applied to these weights, allowing the model to allocate appropriate attention to different parts of the input sequence[51].

This paper demonstrates that Transformers outperform recurrent or convolutional neural networks in translation tasks and can be trained significantly faster. Transformers address several limitations of traditional models, including their ability to learn long-range dependencies in sequence transduction tasks, which is particularly challenging for longer sequences. Unlike RNNs, where each hidden state depends on the previous one, Transformers use self-attention, enabling greater parallelization of computations and reducing the need for sequential operations. This improvement in parallelization contributes to the overall faster training of Transformers compared to traditional models.

3.3 Methodology

3.3.1 Grapheme to Phoneme Conversion

G2P conversion can be considered a machine translation problem where we should translate source graphemes into target phonemes[40]. In linguistics, a phoneme is the smallest unit of spoken sound and is often the one thing that distinguishes one word from another while a grapheme is the smallest unit of a written language whether it carries meaning or corresponds to a single phoneme. The spelling of a word is called a grapheme sequence and the phonetic form is called a phoneme sequence. It is the same process as machine translation based on IPA phonetic transcription as both of them can convert words from their grapheme format to phoneme format[34]. The following diagram(see Fig. 3.4) shows how the G2P model works, where you give it text like "hello world" and it gives you a phonetic transcription. At present, there are three most common types of machine translation: 1)

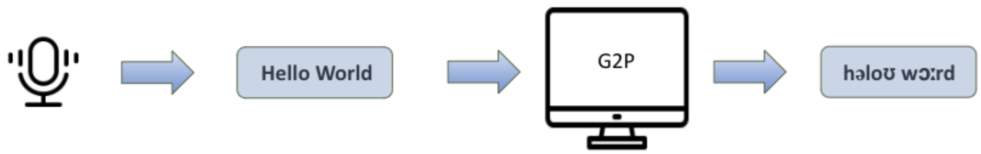


Figure 3.4: Grapheme-to-Phoneme Conversion in APTgt

Rule-based Machine Translation, 2) Statistical Machine Translation, and 3) Neural Machine Translation[55]. Neural machine translation is an end-to-end learning approach for automated translation that applies artificial neural networks to predict the likelihood of word sequences[53, 55]. In this study, we focused on neural machine translation and leverage the Transformer model, the most state-of-art deep learning technology, to build our G2P converter based on IPA symbols.

3.3.2 International Phonetic Alphabet

Multiple IPA keyboards are embedded in the APTgt system to support the phonetic input required for exams or assignments. The APTgt system, being inherently based on the IPA, leverages this widely accepted international standard used by linguists to describe the sounds of spoken language [4]. However, in the context of G2P conversion, the widely used dataset is the Carnegie Mellon University Pronunciation Dictionary (CMUDict) [1, 2, 29]. CMUDict relies on the ARPAbet phoneme set, which consists of 39 phonemes[2]. To further enhance the language support capabilities of APTgt, we aim to build a G2P converter based on IPA dictionaries instead of CMUDict. This approach aligns with the IPA foundation of the APTgt system and opens up possibilities for expanded language support. By utilizing IPA dictionaries, APTgt can accommodate a wider range of languages, remaining consistent with its existing IPA-based framework. Table 3.2 gives the phoneme mapping between IPA notation and the ARPAbet symbol set used in CMUDict[1].

Table 3.2 Mapping between CMUDict's ARPAbet and IPA Symbols

IPA	ARPAbet(CMU)	Examples	IPA	ARPAbet(CMU)	Examples
ɑ	AA	odd	l	L	lee
æ	AE	at	m	M	me
ʌ	AH	hut	n	N	knee
ɔ	AO	ought	ŋ	NG	ping
aʊ	AW	cow	oʊ	OW	oat
aɪ	AY	hide	ɔɪ	OY	toy
b	B	be	p	P	pee
tʃ	CH	cheese	r	R	read
d	D	deer	s	S	sea
ð	DH	thee	ʃ	SH	she
ɛ	EH	Ed	t	T	tea
ɝ	ER	hurt	θ	TH	theta
eɪ	EY	ate	ʊ	UH	hood
f	F	fee	u	UW	two
g	G	green	v	V	vee
h	HH	he	w	W	we
ɪ	IH	it	j	Y	yield
i	IY	eat	z	Z	zoom
dʒ	JH	gee	ʒ	ZH	seizure
k	K	key			

3.3.3 Transformer

The Transformer is a deep-learning model that adopts the self-attention mechanism. Without using any recurrent layers, the self-attention mechanism plays an important role in the Transformer model. The main idea for attention is that it allows the decoder to process the entire input at once and extract the necessary information for subsequent decoding. The Transformer is organized by self-attention and a fully-connected layer for both the encoder and decoder(See Figs. 3.3, 3.5). Each encoder is composed of two major elements:

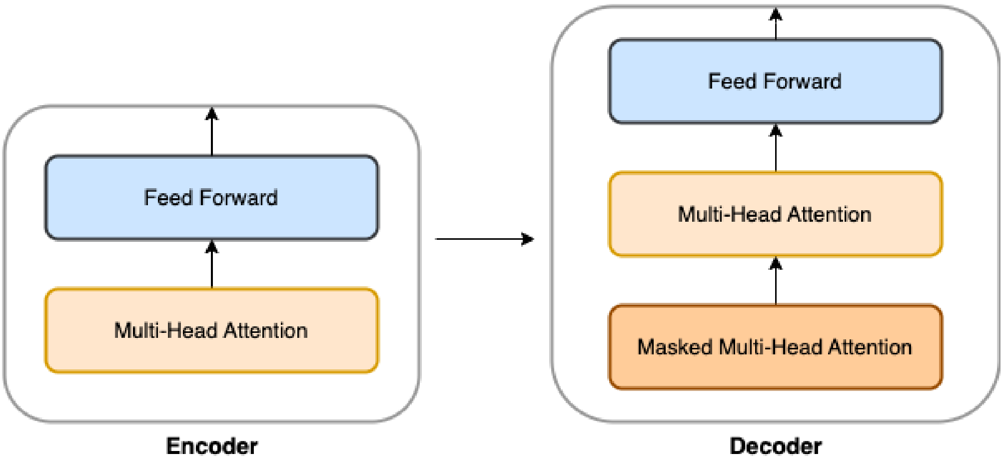


Figure 3.5: The simplified Encoder-Decoder Structure of Transformer Architecture

a multi-head self-attention mechanism and a feed-forward layer. The decoder shares several similarities with the encoder and consists of two multi-head self-attention mechanisms and one feed-forward layer. The encoder maps input sequences into attention-based representations, while the decoder then takes the continuous representations and produces the output

The attention adopted in the transformer is a scale dot-product attention mechanism(see Fig. 3.6), in which the dot products are scaled down by a scaling factor of $\sqrt{d_k}$. Query Q represents a vector word, keys K are all other words in the sequence, and value V represents the vector of the word. The attention function can be represented as formula 3.1[51, 34].

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{3.1}$$

The multi-head attention mechanism is a linear projection of Q, K, V in h times (see Fig. 3.7). The idea of multi-head attention is to compute the scale dot-product attention h

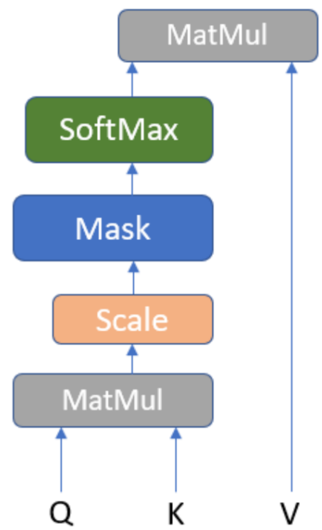


Figure 3.6: Scaled Dot-Product Attention[51]

times in parallel, concatenate the results and project the concatenation to produce the result. Each head of the multi-head attention extracts the specific representation, which allows the whole system to receive information from different representation subspaces[51, 34]. The multi-head attention function:

$$multihead(Q, K, V) = concat(head_1, ..., head_n)W^O \tag{3.2}$$

where

$$head_i = attention(QW_i^Q, KW_i^K, VW_i^V) \tag{3.3}$$

W_i^Q, W_i^K, and W_i^V are the respective weight matrices calculated from Q, K, and V.

3.3.4 Evaluation Metrics

Word Error Rate (WER) is a commonly used metric for evaluating the performance of a machine translation system[10]. It is derived from the Levenshtein distance algorithm and quantifies the percentage of words in which the predicted word sequence differs from

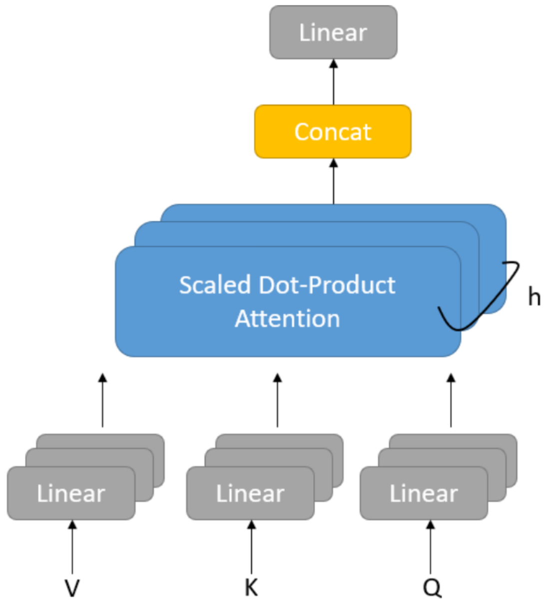


Figure 3.7: Multi-Head Attention[51]

the reference word sequence[30, 56]. WER is calculated by determining the number of word-level substitutions, insertions, and deletions required to align the recognized words with the reference words after aligning them using dynamic string alignment[28]. It can be computed using the following formula(see Equation 3.4):

$$WER = \frac{S + D + I}{S + D + C}, \tag{3.4}$$

Where S, D, and I represent the number of substitutions, deletions, and insertions respectively, and C is the number of correct words.

Phoneme Error Rate (PER) metric evaluates all the mismatches between the predicted and the reference phoneme sequences. Similar to WER, an alignment is performed between the predicted phoneme sequence and the reference phoneme sequence. The alignment determines the minimum number of substitutions, insertions, and deletions needed to align the two sequences, working at the phoneme level instead of the word level.

3.4 Implementation

In this section, we provide a detailed description of the implementation of the Transformer-based G2P converter. We discuss the dataset used, training parameters, programming language and libraries, and hardware resources utilized.

The dataset used for training and evaluating the transformer-based G2P converter was collected from Open-licensed dictionary data available on GitHub. It consists of five languages: English, French, Spanish, Germany, and Dutch. With a meticulous selection, the dataset consists of approximately 56,7000 entries, providing a diverse set of grapheme-to-phoneme mappings. In table 3.3, we provide information about the dataset used for G2P training, including the language, dataset size, training set size, and validation set size.

Table 3.3 Datasets for G2P Training

Language	Total Entries	Training Set	Validation Set
English	125,912	100,730	25,182
French	122,986	98,389	24,597
Spanish	99,315	79,425	19,890
Dutch	**121,199**	**96,959**	**24,240**
German	**98,260**	**78,608**	**19,652**

Preprocessing steps specific to the G2P task were applied to the dataset. These steps included the removal of punctuation and the handling of rare or out-of-vocabulary words. These preprocessing steps ensure that the data is properly formatted and ready for training the transformer-based G2P converter. The dataset was further split into training and validation sets, where 80% of the data was allocated for training the model, and the remaining 20% was for validating the model's performance.

The transformer-based G2P converter was implemented using the PyTorch framework[39]. A set of training parameters was used during the implementation(see Table 3.4). We opted for a six-layer Transformer model with a hidden size of 512 and employed the Adam optimizer during training, setting the learning rate to 0.0001. Additionally, the training process

Table 3.4 G2P Training Parameters

Parameter	Value
Encoder Layers	6
Decoder Layers	6
Number of epochs	120
Batch size	32
Learning rate	0.0001
Dropout rate	0.1

was conducted for 120 epochs, with a batch size of 32 and a dropout rate of 0.1. To ensure efficient model training, the NVIDIA Tesla P100 graphics card was utilized, providing the necessary computational power[34].

3.5 Result & Analysis

In our previous publication[34], we presented the model's performance on three datasets of English, French, and Spanish languages, which serve as a valuable foundation for my current research. The results of our G2P conversion system are summarized in Table 3.5. The table shows that our system achieved a decent conversion accuracy, with an average PER of 2.15% and an average WER of 12.14%. These results indicate that our system is effective in accurately converting graphemes to phonemes for the given dataset and languages[34].

Table 3.5 PER and WER for English-IPA, Spanish-IPA, and French-IPA dataset

Language	PER(%)	WER(%)
English	2.6	11.43
French	2.14	12.7
Spanish	1.7	12.3

Building upon this existing work, this study expands the capability of our G2P converter by incorporating two additional languages Dutch and German. We trained two models for these two languages and the same evaluation metrics of PER and WER were employed to evaluate the accuracy of the conversion.

Training Loss

The training loss curves for the G2P converter model during the 120 epochs of training on Dutch and German languages are demonstrated in Figures 3.8 and 3.9. It shows a gradual decrease in loss, indicating improvements in the system's performance over time.

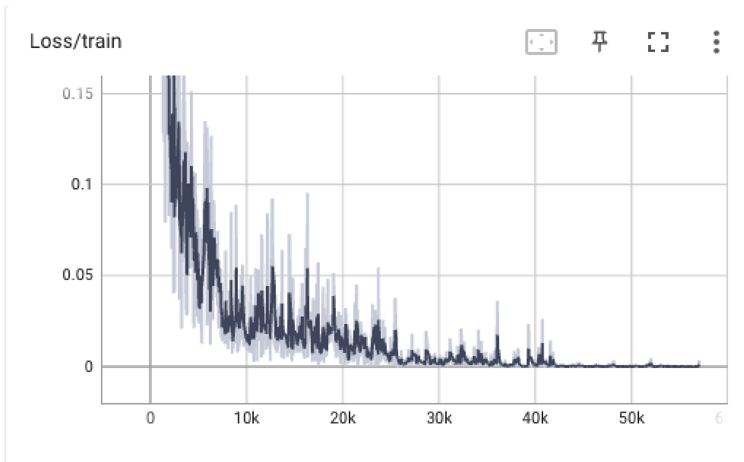


Figure 3.8: Training Loss Curves for Dutch G2P Model (X-axis: Steps, Y-axis: Loss Values)

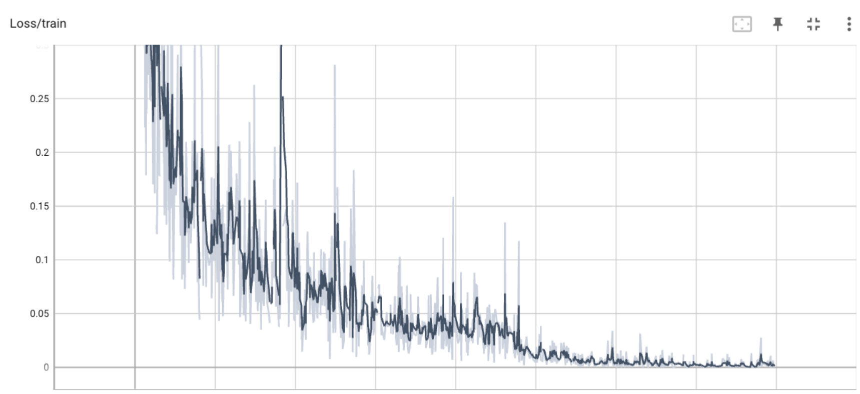
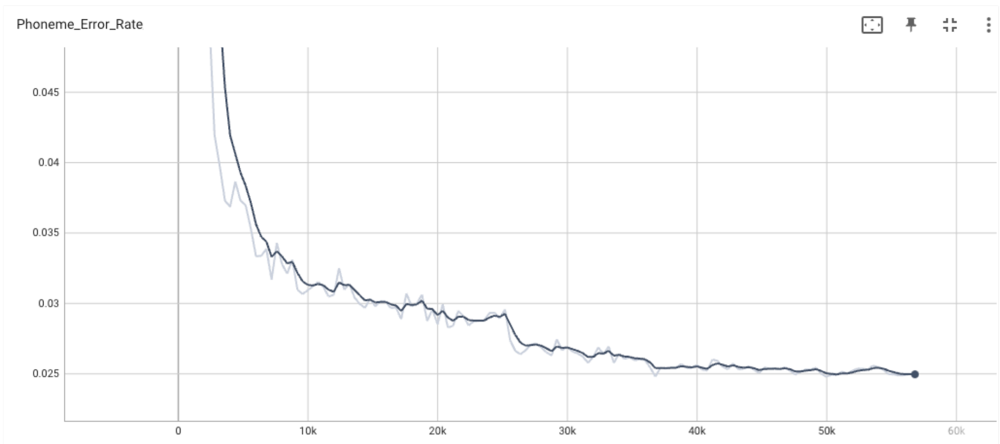


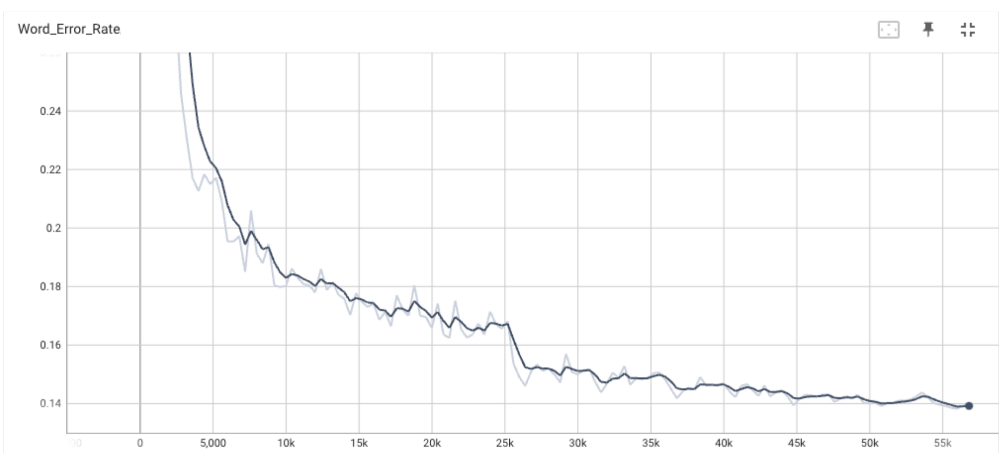
Figure 3.9: Training Loss Curves for German G2P Model (X-axis: Steps, Y-axis: Loss Values)

PER and WER

Figure 3.10 presents the PER and WER performance of our G2P model on the Dutch language, while Figure 3.11 shows the PER and WER performance on the German language. These figures demonstrate the gradual improvement in PER and WER over the training epochs. These reduced low PER and WER values affirm the increased accuracy of our system in converting graphemes to phonemes for multiple languages conversion.



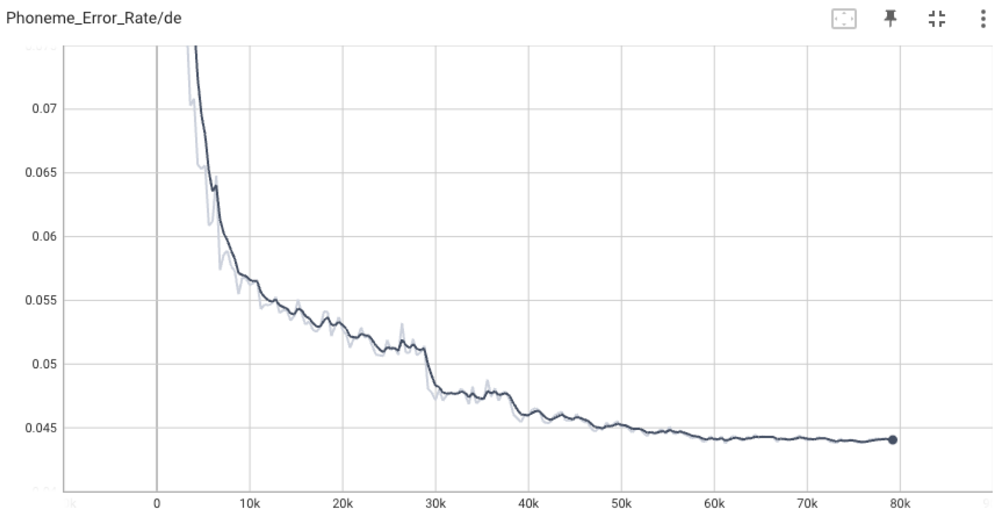
(a) PER Performance of G2P model on Dutch (X-axis: Steps, Y-axis: PER)



(b) WER Performance of G2P model on Dutch (X-axis: Steps, Y-axis: WER)

Figure 3.10: Performance of G2P Model on Dutch-IPA dataset

Table 3.6 presents the PER and WER statistics obtained for each language in our G2P translation system. The Dutch language achieved a PER of 2.49% and a WER of 12.9%. German is known for its complex phonological rules and a relatively large inventory of



(a) PER Performance of G2P model on German (X-axis: Steps, Y-axis: PER)

(b) WER Performance of G2P model on German (X-axis: steps, Y-axis: WER)

Figure 3.11: Performance of G2P Model on German-IPA dataset

phonemes compared to some other languages. This complexity poses challenges for accurate grapheme-to-phoneme tasks[52]. Despite the complexity of the German language, our G2P model achieved a PER of 4.39% and a WER of 26.2%, indicating its effectiveness in capturing the phonetic representations of German words as the conversion accuracy at the phoneme level is over 95%.

Table 3.6 PER and WER for Dutch-IPA, German-IPA dataset

Language	PER(%)	WER(%)
Dutch	2.49	12.9
German	4.39	26.2

The performance compared with the baseline model[46] which is an RNN Seq2Seq model is summarized in Table 3.7. In comparison, our model outperforms the baseline in four languages and performs slightly worse in German. But on average, our model has better performance over these five languages.

Table 3.7: Performance Comparison between Baseline and our Transformer-based Multilingual Model

	Baseline[46]		Our work	
Language	PER(%)	WER(%)	PER(%)	WER(%)
English	12.9	52.8	**2.6**	**11.43**
French	7.7	43.1	**2.14**	**12.7**
Spanish	5.3	36.4	**1.7**	**12.3**
Dutch	2.8	13.5	**2.49**	**12.9**
German	**4.2**	**18.4**	4.39	26.2
Average	6.58	32.84	**2.66**	**15.11**

Translation Samples

Table 3.8 and 3.9 illustrate some samples from the G2P converter model's output for German and Dutch languages. The tables highlight any mismatches, substitutions, and deletions that occur between the model's predictions and the ground truth.

Table 3.8 Translation Samples for Dutch with Ground Truth Comparisons

Dutch Words	Prediction	Ground Truth
aanstelling	anstɛlɪŋ	anstɛlɪŋ
onderscheid	ɔndərsxɛit	ɔndərsxɛit
rijkleding	rɛikledɪŋ	rɛikledɪŋ
verbazingwekkend	vərbazɪŋɛkənt	vərbazɪŋʋɛkənt
bevestigen	bevɛstə	bəvɛstəɣə

Table 3.9 Translation Samples for German with Ground Truth Comparisons

German Words	Prediction	Ground Truth
krümmender	kʁʏməndɐ	kʁʏməndɐ
hemdsärmeligem	hɛmt͡sɛʁməlɪgəm	hɛmt͡sʔɛʁməlɪgəm
ermäßige	ɛɐ̯mɛsɪgə	ɛɐ̯mɛsɪgə
räumet	ʁɔɪmət	ʁɔɪ̯mət
fußballerndes	fusbalɐndəs	fusbalɐndəs

3.6 Conclusion

In this study, our primary objective was to enhance the phonetic exam E-learning system and facilitate linguistic instructors in generating phonetic transcription exams more comfortably and efficiently. To achieve this, we developed a Transformer-based multilingual Grapheme-to-Phoneme converter that demonstrated satisfactory conversion accuracy.

In the future, we plan to conduct additional experiments using the same dataset to investigate the performance differences among various models. Furthermore, we plan to expand the system's language support by incorporating additional languages. By training and fine-tuning the model using more language corpora, we aim to improve the translation performance and overall robustness of the system. This expansion will enable our system to effectively handle multilingual inputs and cater to a wider range of user needs.

Chapter 4

Speech-to-IPA System

A portion of this chapter was published in the proceedings of HCII July 2023 [41]

4.1 Introduction

In this chapter, we propose a novel approach to further optimize the APTgt system for generating linguistic exams and learning resource materials by incorporating an Automatic Speech Recognition (ASR)-based Speech-to-IPA system into APTgt. Building upon the improvements made to the user interface and the implementation of the Grapheme-to-Phoneme (G2P) converter, we aim to leverage ASR technology to overcome the limitations associated with the G2P system and enhance the efficiency and usability of the APTgt system.

4.1.1 Research Problem & Motivation

Although G2P is efficient and cost-effective for creating a large-scale word bank and can bring more convenience to generate exams and resources in APTgt, it has limitations. The system requires prior knowledge of the speech text and is highly dependent on correct spelling, making it sensitive to any changes in the input. Even small misspellings or typos can result in different phonetic transcriptions. Additionally, we found no evidence in the literature that G2P supports disordered speech. Disordered speech often lacks corresponding text, making it challenging to accurately transcribe what individuals with speech disorders are saying, even when given a text to read.

The motivation of this study is to develop a Speech-to-IPA system that operates at the phone level and can convert speech directly into phonetic transcriptions. Compared with the G2P system(see Fig. 3.4), this module eliminates the need for text mediation(see Fig. 4.1).

The system is designed to transcribe both regular and disordered speech and to generate

Figure 4.1: Speech-to-IPA module in APTgt

phonetic learning resources automatically. The study also aims to expand the word bank for disordered speech and provide linguistic instructors with an efficient tool for generating and grading phonetic transcription exams automatically.

4.1.2 Research Questions

The research questions that this study is set to answer are as follows:

1. Recognition of child speech and disordered speech are data-scarce tasks. What dataset will we use to train and test child speech and disordered speech?

2. What kind of data features shall we use for speech recognition?

3. How to characterize the data to better observe the result for the different populations?

4. How to measure the performance of the recognizer and what performance it can achieve?

4.1.3 Research Hypothesis

- H1: The use of a Speech-to-IPA system will make it easier and faster to create phonetic exam resources for the E-learning system.

- H2: The Speech-to-IPA system will reduce the need for manual transcription and user involvement to create learning resources.

- H3: The Speech-to-IPA system will be able to better recognize disordered speech versus the traditional ASR system.

To address the research questions outlined, we propose utilizing a phone-level ASR system that applies Mel-frequency cepstral coefficients (MFCCs) as features[54]. The system is based on a deep learning Sequence to Sequence model that utilizes bidirectional Long Short-Term Memory (LSTM) as the model architecture[23, 19]. For training and testing the ASR system, we incorporate the Speech Exemplar and Evaluation Database (SEED) [7] and TORGO Dysarthria Speech Database[42]. These datasets serve as valuable resources as they encompass both disordered and non-disordered speech samples from adults and children. Additionally, the TORGO dataset includes phonetically transcribed disordered speech data specifically from adults.

The proposed speech-to-IPA module enables linguistic instructors to generate numerous phonetic transcription exam resources with minimal manual intervention, making our phonetic E-learning system more efficient and intelligent. The main contribution is the integration of machine learning technology, which reinforces the phonetic tools to better serve instructors with exam and training resource creation. Additionally, the study will make research in communication disorders more tractable by expanding the system's ability to transcribe disordered speech.

4.2 Literature Review

Automatic Speech Recognition: A Shifted Role in Early Speech Intervention? [21]

Automatic speech recognition (ASR) refers to a series of techniques combining signal processing, statistical modeling, and machine learning to interpret human speech typically by deciphering input acoustic signals into phones or other linguistic elements such as syllables, words, or phrases. The potential of ASR to support computer-based tools to improve the efficacy of the traditional face-to-face clinician-client dyad and the potential to provide new modes of intervention, outside of face-to-face sessions with an SLP has been recognized previously.

In this article, Hamidi and Baljko review a number of extant systems that employ ASR for speech and found that these ASR-based systems that support speech training for children face significant challenges in improving performance on non-standard speech and designing effective feedback. The empirical qualitative data showed that children liked the idea of playing with a computer and getting explicit feedback but found the visual feedback confusing and unhelpful. And the child users suggested adding more game-like features, such as goals and rewards, to make it more engaging. Some systems, such as VocSyl, focus on engagement and motivation, and visualizations help SLPs demonstrate specific aspects of vocalization. However, the results showed that in these applications, if corrective feedback is given in the absence of SLPs or parents to facilitate their interpretations, children were less motivated to continue using their speeches.

The authors conclude that although ASR is challenged by certain design requirements, it supports the requirement that the system is engaging, and interactive, and motivates repeated speech productions by the child. They recognize the limitation of ASR to analyze non-standard speech but their fieldwork indicates that it can be effective to subordinate the accuracy of ASR to its use as a facilitator and encourager of speech interaction.

An Efficient MFCC Extraction Method in Speech Recognition[22]

Automatic speech recognition (ASR) by machines has been studied for decades. The first step in any ASR system is to extract features of audio signals. Among the different kinds of parametric representations for acoustic signals, Mel-Frequency Cepstrum Coefficients (MFCC) are the most widely used in ASR systems. Conventional MFCC extraction algorithm involves the following implementation steps, including pre-emphasis, windowing, Fast Fourier Transform (FFT), Mel-Frequency filter bank, logged energy, and delta calculations. A total of over 1,700 multiplications are required for each speech segment during the above-mentioned steps. As a result, this algorithm requires a huge amount of calculations which increase the cost and degrade the performance of the hardware recognizer.

In this paper, Han et al. proposed a new and efficient algorithm for extracting MFCC for speech recognition that only requires half of the multiplication steps. First of all, the complex multiplication operation in the pre-emphasis step is replaced with simple addition and shift operations without affecting the recognition accuracy. Then the overlap function, which was originally combined with the window function in the conventional approach, is separated and moved after the filter bank. In this new design, input speech is divided into short segments called subframes instead of overlapping frames, and one subframe consists of 80 points with no overlap between them. Therefore, the length of the Hamming window can be reduced from 160 points to 80 points, and the amount of computation in the FFT is cut in half due to the new window size. Furthermore, the authors modify the filter bank from equally spaced triangular filters to equally spaced rectangular filters. The simulation results indicate that 23 equally spaced rectangular filters produce the highest recognition accuracy. In addition, moving the new overlap operation to the end of the spectral calculation helps to reduce the computation in half, thanks to the benefits of the previous modification steps. In brief, Han introduces a new extraction algorithm and demonstrates that the proposed algorithm reduces the number of multiplications from 1708 to 804, while the recognition accuracy drops by only 1.5%. The new algorithm is more efficient than the original algorithm in hardware implementation.

Dysarthric Speech Recognition using Convolutional Recurrent Neural Networks[11]

In this paper, a deep architecture of the Convolutional Recurrent Neural Network (CRNN) model was developed and compared with the Vanilla Convolutional Neural Network (CNN) model in terms of performance. Both models were trained using samples from the Torgo dataset, which includes a mixture of disordered and non-disordered speech data. The experimental findings demonstrate that the CRNN model achieved an accuracy of 40.6%, outperforming the Vanilla CNN model, which achieved an accuracy of 31.4%. The proposed

CRNN model achieved a 9 percent improvement in recognition accuracy, indicating the effectiveness of the proposed hybrid structure of the CRNN in improving the recognition of disordered speech. It serves as the baseline model for assessing our work.

Automatic Speech Recognition of Disordered Speech: Personalized models outperforming human listeners on short phrases[20]

Automatic Speech Recognition technologies have the potential to help individuals with Speech impairments by facilitating more real-time conversations through machine translation. However, while ASR accuracy has improved significantly over recent years due to the increased computational power of deep learning systems and the availability of large training datasets, disordered speech recognition is still unacceptably low, rendering the technology unusable for speakers who could benefit the most. The poor recognition is partly due to the complexity of atypical speech patterns and insufficient training data. To address these challenges, Green et al. focused on using ASR models personalized to the disordered speech of their 432 participants rather than more generalizable speaker-independent models. In their study, the full dataset contained recordings collected from 432 speakers with various speech impairment types and severities. It was split into three subsets, which are the High or low Word Error Rate(WER) subset, the Surprisingly High WER subset, and the Human transcription WER subset, to identify the factors associated with ASR performance and compare the accuracies of personalized models with those of human listeners. In terms of the models, the first speaker-independent ASR model (SI-1) was accessed via Google's Speech-to-Text API, and the second speaker-independent model (SI-2) was an end-to-end ASR model based on the RNN-T architecture. The encoder network and predictor network in SI-2 consist of 8 layers and 2 layers of uni-directional LSTM cells respectively. Inputs were 80-dimensional log-Mel filterbank energies. Outputs were probability distributions over a 4k word piece model vocabulary. For the personalized ASR model, the researchers conducted the optimized fine-tuning procedure on their SI-2 model and applied SpecAugment as a regularization method.

Compared to the two speaker-independent ASR models in this paper with median WERs of 31.5% and 29.4%, the accuracy of the proposed personalized models was excellent (i.e., median WER of 4.8%) for most speakers and similar to or better than those of expert human listeners. And this approach leads to highly accurate models that can achieve up to 85% improvement in the word error rate in disordered speech compared to out-of-the-box speech models trained on typical speech. This result demonstrates the efficacy of personalized ASR models for recognizing a wide range of speech impairments and severities, with the potential for making ASR available to a broader population of users.

End-to-end acoustic modeling for phone recognition of young readers[18]

Automatic recognition systems for child speech are lagging behind those dedicated to adult speech in the race of performance. This phenomenon is due to the high acoustic and linguistic variability present in child speech caused by their body development, as well as the lack of available child speech data. Young readers' speech additionally displays peculiarities, such as slow reading rate and the presence of reading mistakes, that hardens the task.

The work in this paper attempts to tackle the main challenges in phone acoustic modeling for young child speech with limited data and improve understanding of the strengths and weaknesses of a wide selection of model architectures in this domain. By comparing recent end-to-end models such as RNN, LAS, and Transformer to a baseline hybrid DNN-HMM model for phone recognition, the researchers find that transfer learning techniques are highly efficient on end-to-end architectures for adult-to-child adaptation with a small amount of child speech data. Through transfer learning, a Transformer model complemented with a Connectionist Temporal Classification (CTC) objective function, reaches a phone error rate of 28.1%, outperforming a state-of-the-art DNN-HMM model by 6.6% relative, as well as other end-to-end architectures by more than 8.5% relative. An analysis of the models' performance on two specific reading tasks (isolated words and sentences) is provided, showing the influence of the utterance length on attention-based and CTC-based models. In summary, the Transformer+CTC model displays an ability to better detect reading mistakes made by children, which can be attributed to the CTC objective function effectively constraining the attention mechanisms to be monotonic.

4.3 Methodology

The methodology employed in this study focuses on developing a Speech-to-IPA module for APTgt, which enables instructors to bypass the text mediation in the g2p system and generate phonetic transcriptions by directly uploading disordered speech samples or speaking words or sentences into a microphone. The module aims to analyze and process audio signals, extract relevant features, and detect phones based on these features.

To achieve this, we utilized MFCCs as the representation of the sound features. MFCCs are commonly used in speech-processing tasks due to their effectiveness in capturing acoustic characteristics[54]. Additionally, we employed a bidirectional LSTM as an encoder within the Seq2seq model. LSTMs are known for their ability to model temporal dependencies and capture long-range dependencies in sequential data[23]. For the training and testing dataset, we utilized the SEED dataset, which includes speech samples from adults and children.

This section will further elaborate on the data preparation, feature extraction, and model architecture. These details will provide a comprehensive understanding of how the Speech-to-IPA module was developed and validated within the APTgt system.

4.3.1 Data Preparation

We will use SEED as one of our datasets which was created for clinical training in articulatory phonetics and speech science(see Fig. 4.2). The SEED contains about 17,000 high-quality recorded speech samples along with their text, grouped by age (child vs. adult) and speech health status(with or without speech disorder)[47].

The data preparation process on the SEED dataset involves several steps to ensure the quality and compatibility of the training data. It includes filtering out speech samples that do not meet certain criteria. Samples with low volume were excluded from the dataset as they

Figure 4.2: The Speech Exemplars and Evaluation Database[9]

can impede the accurate detection and processing of the speech signal, thereby potentially compromising the overall accuracy of the ASR system. Additionally, the speech samples are standardized to a mono-channel WAV format with a bit depth of 16, ensuring uniformity and optimal compatibility throughout the training process.

Another important preprocessing step for the SEED data is the annotation of the speech data for speech-to-IPA conversion. To achieve this, we utilized our G2P converter to translate the English text in the SEED corpus into its corresponding IPA phonetic forms. For example, the word "*impossible*" should be converted to "/ɪmpɑsəbəl/". Subsequently, we manually inspected the result to ensure all samples and phonetic transcriptions matched up correctly.

For our training and testing of disordered speech, we utilized the TORGO database, a well-known dysarthric speech database that contains aligned acoustic and articulatory recordings from 15 speakers[42]. Dysarthria, as defined by the American Speech-Language-Hearing Association (ASHA), is a common speech disorder caused by muscle problems[3]. Within the TORGO dataset, eight out of the 15 speakers (5 males, 3 females) have dysarthria, while the remaining seven speakers (4 males, 3 females) serve as control subjects without any speech disorders[25, 24]. The severity level of the speech disorder for each of the eight dysarthric speakers was evaluated by a speech-language pathologist[42, 24].

The database comprises recordings of various speech elements, including single words, sentences, and descriptions of photograph contents provided by the speakers. The single words encompass English digits, international radio alphabets, the twenty most frequent words in the British National Corpus (BNC), as well as a set of words selected by Kent et al. [25] to account for relevant phonetic contrasts. The sentences were sourced from the Yorkston-Beukelman assessment of intelligibility [57] and the TIMIT database [17]. Additionally, to incorporate dictation-style speech, subjects were asked to describe the contents of several photographs in their own words [24]. The recordings were captured using two types of microphones: an array microphone and a head-mounted microphone. Overall, approximately three hours of speech were recorded across multiple sessions [24]. The specific number of recordings from speakers with speech disorders in the TORGO Corpus is outlined in Table 4.1.

Table 4.1 Details of TORGO Disorder Speech Database(*F: female speaker. M: male speaker. S-M represents severe-moderate category of dysarthria patients)[24]

Speaker	F01	M01	M02	M04	M05	F03	F04	M03	Total
Disorder	Severe	Severe	Severe	Severe	S-M	Moderate	Mild	Mild	-
#Utterance	228	739	772	659	601	1097	675	806	5,577

In the preprocessing step on the TORGO dataset, we performed a relabeling process on the phonetic transcriptions of the selected samples. Specifically, we converted the phonetic transcriptions from the ARPAbet phone set to the IPA set using mapping rules (refer to Table 3.2) outlined in Section 3.3.2. For instance, consider the disordered speech of speaker "F03" with the given English text "double". The corresponding phonetic transcription was "d ah b ah l". We relabeled it to "d ʌ b ʌ l" based on the mapping rule. This relabeling process enables us to develop and evaluate our speech-to-IPA conversion models.

4.3.2 Mel-frequency Cepstral Coefficients

MFCCs are the most widely used parametric representations for acoustic signals in ASR systems[12]. The MFCCs extraction algorithm involves the following implementation steps:

1. Pre-emphasis increases the magnitude of energy at higher frequencies.

2. Split the signal into short frames.

3. For each frame, apply the Fast Fourier Transform (FFT) to convert the signal from the time domain to the frequency domain. Calculate the power spectrum of each frame using the following equation(refer to Eq. 4.1):

$$P = \frac{|FFT(X_i)|^2}{512} \tag{4.1}$$

4. Apply Mel-scale filterbanks to the power spectrum of the signal and take the logarithm of all filter bank energies. The Mel-scale maps the actual frequency to the frequency that human beings perceive. The formula for the mapping is(refer to Eq. 4.2):

$$Mel(f) = 2595log(1 + \frac{f}{700}) \tag{4.2}$$

5. The MFCCs are extracted after applying the Discrete Cosine Transform (DCT).

4.3.3 Bidirectional Recurrent Neural Network

A recurrent neural network (RNN) is a type of neural network commonly used in speech recognition. The network consists of an input layer, a hidden layer, and an output layer, where each output layer unit has a feedback connection to itself. The feedback loops remember historical inputs which allows them to make decisions by considering current inputs while learning from previous inputs[12]. In this way, RNNs can gain a deeper understanding of the sequence and its context than other types of deep learning algorithms, enabling more precise prediction results.

LSTM is an RNN architecture used in ASR systems. It contains special units called memory blocks in the recurrent hidden layer and is better for maintaining long-range connections, recognizing the relationship between values at the beginning and end of a sequence[43].

Bidirectional LSTM adds one more LSTM layer, which reverses the direction of the information flow. The architecture of a one-layer bidirectional LSTM network is illustrated in Figure 4.3. Unlike standard LSTM, the input flows in both directions and is capable of utilizing information from both sides.

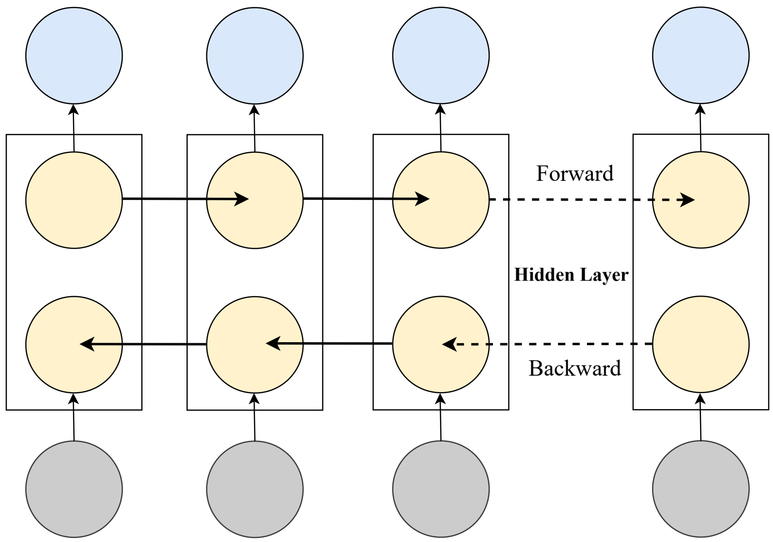


Figure 4.3: Bidirectional LSTM network architecture

4.4 Implementation

The ASR-based Speech-to-IPA module in the system will include an Encoder-Decoder Seq2Seq Model. The overview of the training procedure is demonstrated in Figure 4.4. In our experiment, the model applies 40-dimensional MFCCs as features that carry the

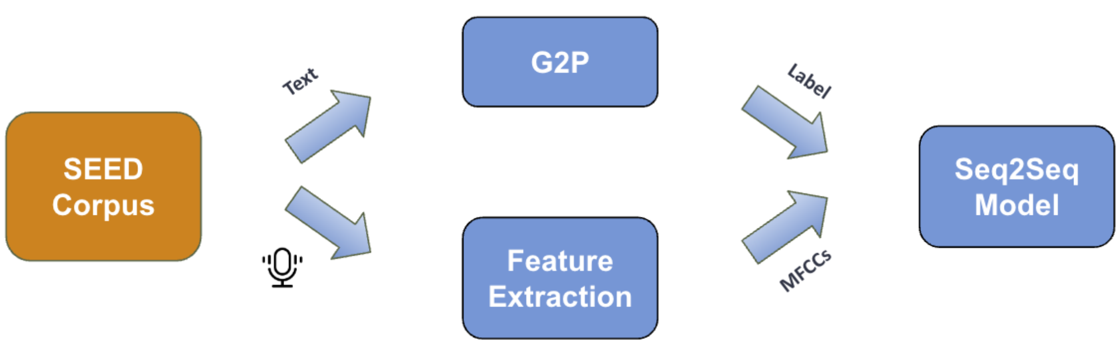


Figure 4.4: The raining procedure of the Seq2Seq Model in Speech-to-IPA system

information we can use to detect phones in speech. The Bidirectional LSTM (BiLSTM) network with five hidden layers and 1024 hidden units at each hidden layer is trained on the MFCC features. The BiLSTM encoder is built with PyTorch and the training is done by Stochastic Gradient Descent with an initial learning rate of 0.01. The frame batch size was set as 6000, and the training was conducted for 100 epochs. This network gives the probability of each phone in the inventory for each sound. Subsequently, the decoder finds the most probable symbols from the phone inventory based on the probability values and outputs the recognized phonetic symbols.

The model evaluated on a total of 10131 speech samples in SEED, of which approximately 30% are words and about 70% are sentences(refer to Table 4.2). 95% of the selected samples are used for training and validation, and the remaining data are used for testing.

Table 4.2 SEED Corpus Size in Utterances for Training and Testing

Dataset	Adult	Children	Total
Sentence SEED	2,054	1,202	3,256
Word SEED	3,467	3,408	6,875
Full SEED	5,521	4,610	**10,131**

In our experiments on disordered speech, we exclusively focused on the recordings of disordered speech in the TORGO database from the head-mounted microphone, which captured audio at a sampling rate of 16 kHz. This sample rate aligned with our selection criteria for the SEED database. However, the recordings of the male speaker "M03" were excluded from our training and test sets due to the unavailability of phonetic transcriptions for his speech. The details of the disordered speech data in TORGO used for training and testing are illustrated in Table 4.3. We followed the same split ratio as the SEED dataset, where 95% of the selected samples were used for training and validation, and the remaining 5% were used for testing.

We initiated our implementation by employing transfer learning, utilizing the model trained on the SEED dataset, which is over four times larger than the TORGO dataset

Table 4.3 TORGO Corpus Size in Utterances for Training and Testing (*F: female speaker. M: male speaker. S-M represents a severe-moderate category of dysarthria patients)

Speaker	F01	M01	M02	M04	M05	F03	F04	Total
Disorder	Severe	Severe	Severe	Severe	S-M	Moderate	Mild	-
#Utterance	134	386	409	424	523	577	250	2,703

selected for our research. Next, we performed finetuning on the pre-trained model using the entire disordered speech dataset. Moreover, we acknowledged the significance of individual variations among speakers, considering their different severity levels and patterns of errors in their speech. To address these variations, we implemented personalized models for each speaker in our dataset. This process involved fine-tuning the pre-trained SEED model using the disordered speech data from each individual, enabling us to capture their unique speech characteristics and further enhance the accuracy of recognition.

4.5 Result & Analysis

With the implementation of a Speech-to-IPA model for disordered speech, our goal is not to reconstruct and correct words based on detected phones but to transcribe what exactly the speaker has pronounced, including potential phone-level speech errors in disordered speech. Therefore, instead of using the classic WER to measure performance, we used the Phone Error Rate(PER), which is similar to Phoneme Error Rate mentioned in Section 3.3.4. The PER metric considers all mismatches between the recognizer hypothesis and the manual phone-level annotated reference (see definition in Equation 2), with C, I, S, and D respectively, referring to the number of correct detections, insertions, substitutions, and deletions[18].

$$PER = \frac{I + S + D}{C + S + D} \tag{4.3}$$

4.5.1 Model Performance on SEED dataset

In this section, we present the performance of our models on different categories of the SEED dataset, including sentences and words, as well as the performance on adult and children's speech. We also provide an overall PER for the full SEED dataset. The training loss curve (see Fig. 4.5) demonstrates a decreasing trend, indicating that the model improves its performance and learning from the data. Similarly, the training PER (see Fig. 4.6) also

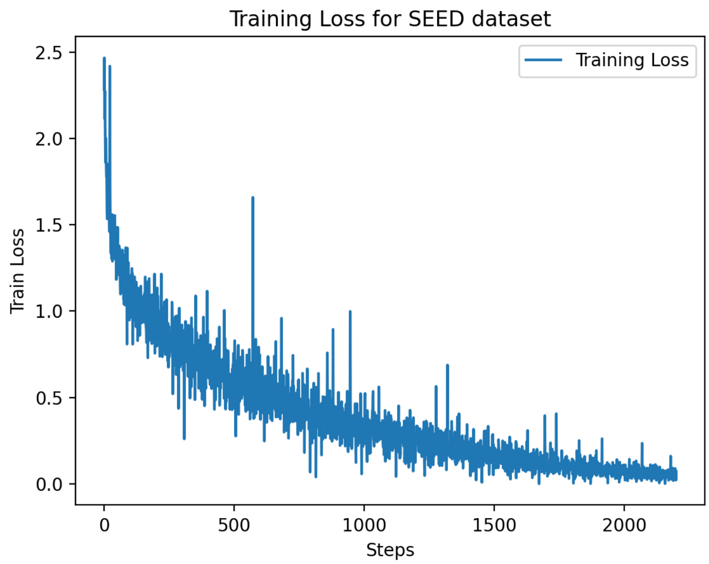


Figure 4.5: Training Loss for SEED dataset

shows a decreasing trend, indicating the model is reducing errors in predicting phonetic transcriptions and improving its overall performance. The statistical results are summarized in Table 4.4.

Table 4.4: Results of model's phone error rate performance on SEED

Dataset	**PER(%)**		
	Adult	**Children**	**Overall**
Sentence SEED	6.02	35.42	16.08
Word SEED	14.45	29.68	26.1
Full SEED	8.55	32.15	**18.56**

The resultant model achieved a PER of 6.02% for adult sentences, 35.42% for children's sentences, and an overall PER of 16.08% on the sentence SEED dataset. For the word SEED dataset, the model achieved a PER of 14.45% for adult words, 29.68% for children words, and an overall PER of 26.10%. Finally, on the full SEED dataset, which combines

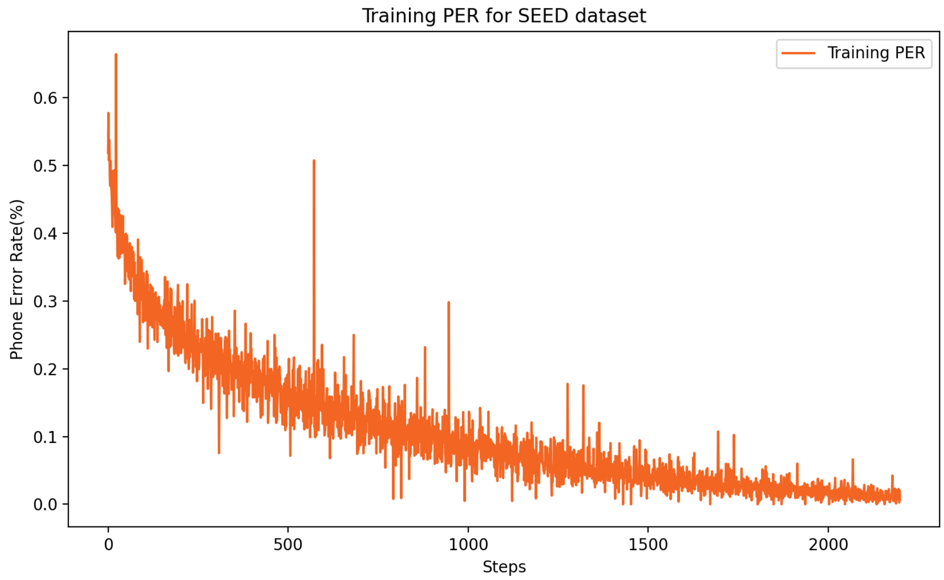


Figure 4.6: Training PER for SEED dataset

both sentences and words, the model achieved a PER of 8.55% for adult speech, 32.15% for children's speech, and an overall PER of 18.56% which means the overall recognition accuracy is 81.44%.

These results indicate that the model performs better on adult speech compared to children's speech across all categories. Additionally, the model achieved the lowest overall PER on the sentence SEED dataset, suggesting that it performs better on sentence-level tasks compared to word-level tasks. This set of results can be considered the benchmark performance on the SEED dataset. It provides a baseline for evaluating the performance of future models and techniques on this dataset.

4.5.2 Model Performance on TORGO dataset

We further evaluated the performance of our models on the disordered speech data from the TORGO dataset. Table 4.5 provides a comparison of the PER between the pre-trained SEED model and the fine-tuned model on the TORGO dataset. The table includes the PER for different severity levels (Severe, S-M, Moderate, Mild) for individual speakers (F01, M01, M02, M04, M05, F03, F04).

Table 4.5 Comparison of %PER between Pre-Trained SEED model and Fine-Tuned Model on TORGO Dataset

Method	Severe				S-M	Moderate	Mild
	F01	M01	M02	M04	M05	F03	F04
SEED model	64.26	74.62	66.86	63.97	64.93	56.99	47.07
Finetuned	51.53	59.09	51.97	46.86	45.15	40.17	26.93
PERreduced	12.73	15.53	14.89	17.11	19.78	16.82	20.14

The results demonstrate that the fine-tuned model outperforms the pre-trained SEED model across all severity levels. The PER is consistently reduced for all speakers after fine-tuning, indicating improved accuracy in phonetic transcription. On average, the PER is reduced by 15.07% for the Severe category, 19.78% for the S-M category, 16.82% for the Moderate category, and 20.14% for the Mild category. By leveraging the fine-tuned model, we are able to achieve enhanced accuracy in converting speech to IPA-based phonetic representations.

The results in Table 4.6 present a comparison of PER between the fine-tuned models and personalized models on the TORGO dataset. The personalized models show even better performance compared to the fine-tuned models across all severity levels.

Table 4.6 Comparison of %PER between Fine-Tuned and Personalized Models on TORGO Dataset

Method	Severe				S-M	Moderate	Mild
	F01	M01	M02	M04	M05	F03	F04
Finetuned	51.53	59.09	51.97	46.86	45.15	40.17	26.93
Personalized	36.85	36.87	27.53	31.63	25.22	27.21	15.58
PERreduced	14.68	22.22	24.44	15.23	19.93	12.96	11.35

The overall average PER for all speakers in the TORGO dataset is 28.7%, indicating a phone-level recognition accuracy of **71.3%**. In comparison, the baseline Convolutional Recurrent Neural Network (CRNN) model trained on the same dataset achieved an accuracy of 40.6%[11]. Our proposed method surpasses the baseline model by **31.3%**, demonstrating its superior performance in accurately transcribing disordered speech.

Specifically, the personalized models achieve an average PER reduction of 19.14% for the Severe category, 19.93% for the S-M category, 12.96% for the Moderate category, and 11.35% for the Mild category when compared to the fine-tuned models.

These reductions in PER demonstrate the effectiveness of personalizing the models to the unique characteristics of each speaker's disordered speech. By tailoring the models to individual speakers, we further improved the accuracy in converting disordered speech to IPA-based phonetic representations.

4.5.3 Recognition Samples

We have generated recognition samples for a sentence and a word from an adult speaker and a child speaker. The unique sample IDs in the SEED dataset are as follows:

- Sentence from adult: 2AU203-11NF44-BIT01
- Sentence from child: 2AU201-24NM3_2-BIT01
- Word from adult: 2AU203-03NM22-MSN04
- Word from child: 2AU201-30NF4_8-MSN04

Table 4.7 displays the prompts of the speech samples, along with the predictions from our model trained on SEED data, and the corresponding ground truth for comparison. The predictions for the adult speaker show no errors, which is expected given the high recognition accuracy of our model on adult speech. However, a few errors were observed in the predictions for the child's speech.

In Table 4.8, we present examples of the Speech-to-IPA model's output for disordered speech in the TORGO dataset. The selected samples cover different severity levels of the disorder. The table illustrates any mismatch, substitutions, and deletions observed between

Table 4.7 Recognition Samples for Speech in SEED dataset with Ground Truth Comparisons.

SEED data	Sentence	Word
Prompt	The baby falls.	Wagon
Ground Truth	ð ə b e ɪ b i f ɔ l z	w æ g ə n
Prediction(Adult)	ð ə b e ɪ b i f ɔ l z	w æ g ə n
Prediction(Child)	ð ə b e ɪ i f ɔ r	w æ k ə n

the model's recognition and the ground truth. The errors show a decreasing trend from the mild severity level to the severe level.

Table 4.8 Recognition Samples for Disordered Speech with Ground Truth Comparisons. *Speech Prompt: *"They carried me off on the stretcher."*. S1 in the speech ID represents the recording session 1.

Speech ID	Severity level	Prediction	Ground Truth
F04_S2_0057	Mild	ð e ɪ k æ r i d m i ɔ f ɑ n ð ʌ s t r ɛ t͡ʃ ɚ	ð e͡ɪ k æ r i d m i ɔ f ɑ n ð ʌ s t r ɛ t͡ʃ ɚ
F03_S1_0196	Moderate	ð e j k æ r i m i ɔ f ɑ n ð ʌ s t r ɛ t͡ʃ ɚ	ð e͡ɪ k æ r i d m i ɔ f ɑ n ð ʌ s t r ɛ t͡ʃ ɚ
M01_S2_0132	Severe	ð e ɪ k æ r i d m i ɔ f ʌ n ʌ n ð ʌ s t ɛ t͡ʃ ʌ	ð e͡ɪ k æ r i d m i ɔ f ɪ n ʌ n d ʌ s t r ɛ t͡ʃ ɚ

As we have trained separate models for the SEED and TORGO datasets, our process can auto-generate all the speech samples in these two databases. The total number of words/sentences that can be auto-generated from the SEED database and TORGO database is approximately 13,000 (refer to Table 4.2 and Table 4.3). Based on the test on a small subset of 217 speech samples, the total execution time for auto-generation is about 93.9 seconds(see Figure 4.7). This indicates that the average recognition time for each speech sample is only 0.43 seconds.

```
Recognition in progress: 100%|                    | 217/217
Execution time: 93.90064907073975 seconds
```

Figure 4.7: Execution Time for Recognition on Subset of 217 Speech Samples

4.6 Conclusion

This speech-to-IPA module can be initially incorporated into APTgt and serves as an auxiliary tool for automatically generating training and examination resources in the field of phonetic transcription, but it has great potential for various applications beyond just E-learning. It can bypass text mediation and directly converts the audio speech signal into IPA symbols, which is helpful in the research of communication disorders where understanding the exact errors of speech is critical.

In this work, we focus on the LSTM-based model with only the MFCCs feature as the first step in identifying phones in speech samples from the SEED dataset. We utilized the SEED and TORGO databases for training and testing and conducted various experiments to evaluate the model's performance. The results showed that fine-tuning the pre-trained model on disordered speech data led to significant improvements in Speech-to-IPA accuracy. Additionally, personalized models tailored to individual speakers further enhanced recognition accuracy, capturing the unique characteristics and patterns of their disordered speech. Our work has established benchmark performance in different categories on the SEED dataset and highlights the accuracy improvements achieved through fine-tuned and personalized modeling approaches for Speech-to-IPA recognition in disordered speech.

Future work will explore other deep architectures and combine additional features with MFCCs to improve accuracy. In addition, we will finetune the model on multiple datasets to enhance its robustness.

Chapter 5

Conclusion

The three studies conducted as part of this research aimed to enhance the Automated Phonetic Transcription Grading Tool (APTgt) and improve its usability and functionality for communication disorders faculty.

Study 1 focused on optimizing the user interface and user experience of APTgt, addressing design and usability concerns identified in previous iterations. The findings highlighted the importance of user-centered design in developing effective online educational tools for linguistic instructors. The updates made in APTgt 2.0, based on the feedback from linguistic professionals and usability experts, have contributed to an improved and efficient tool for phonetic E-learning. In Study 2, a Transformer-based multilingual Grapheme-to-Phoneme converter was developed to enhance the phonetic exam E-learning system. The converter demonstrated satisfactory conversion accuracy and has the potential to support more languages. Study 3 introduced a speech-to-IPA module that directly converts audio speech signals into IPA symbols. This module demonstrated superior recognition accuracy on disordered speech compared to the baseline ASR model. It has the potential to facilitate research in communication disorders by accurately identifying patterns of errors in speech at the phone level.

In summary, the proposed G2P converter can help linguistics instructors to do text-to-IPA tasks with decent conversion accuracy. Meanwhile, the speech- to-IPA module can enable recognize both typical speech and disordered speech at the phone level with improved performance compared with traditional ASR. These allow linguistic instructors to utilize the auto exam generator in APTgt so there will be less need for manual transcription to create the word bank, which is crucial data for the auto exam generator. The word bank can now expand from hundreds of entries to potentially over five thousand or more requiring minimum involvement from linguistic professionals.

Our research has significantly contributed to the enhancement of APTgt, providing linguistic instructors with improved tools for phonetic transcription training and exam generation. The findings emphasize the importance of user-centered design, multilingual support, and automated solutions in the field of phonetic E-learning. Our research's novelty lies in applying deep learning for speech-to-IPA in recognizing disordered speech, specifically for the E-learning domain. This is a challenging task due to the variability in speech patterns and pronunciation problems in disordered speech. Our approach provides the use of speech recognition technology that goes beyond traditional speech recognition and offers benefits for E-learning platforms. Our research will establish a benchmark for Speech-to-IPA on disordered speech, as there is currently limited literature on this topic. By developing and evaluating our approach to recognizing disordered speech, our study can improve the overall effectiveness of the phonetic E-learning system and serve as a reference point for future studies in the context of phonetic E-learning.

There are some limitations and potential areas for future research to explore. As a next step, we plan to apply the module to recognize disordered speech from children once the SEED dataset is fully phonetically transcribed. While the SEED dataset has been a valuable resource for our research, it is worth noting that it contains fewer speech samples from children compared to adults. In future research, it would be beneficial to explore additional child speech datasets, such as the CHILDES database from the Child Language Banks in the TalkBank project[38, 37, 36] which contains corpora with speech samples of children. This expansion can greatly enhance our understanding of speech disorders in children and further strengthen the capabilities of our E-learning system in phonetic transcription. Another limitation is that clinical speech-language pathologists typically transcribe disordered speech with diacritics(i.e. for this work we included the basic IPA set as seen in Figure 2.2 excluding the diacritics ṇ, ṃ, ḷ), which provide additional detail not attempted in our automated system. Future work could aim to incorporate diacritics in automated transcription to capture a more precise representation of disordered speech.

Bibliography

[1] Arpabet. https://en.wikipedia.org/wiki/ARPABET. Accessed: 2023-06-06.

[2] The cmu pronouncing dictionary. http://www.speech.cs.cmu.edu/cgi-bin/cmudict. Accessed: 2023-06-5.

[3] Dysarthria. https://www.asha.org/public/speech/disorders/childsandl. Accessed: 2023-07-07.

[4] International phonetic alphabet. https://en.wikipedia.org/wiki/International_Phonetic_Alphabet. Accessed: 2023-06-06.

[5] Katalon studio. https://en.wikipedia.org/wiki/Katalon_Studio. Accessed: 2023-01-11.

[6] Moodle. https://moodle.com. Accessed: 2023-07-13.

[7] Phonetic transcription. https://en.wikipedia.org/wiki/Phonetic_transcription. Accessed: 2022-12-14.

[8] Scrum (software development). https://en.wikipedia.org/wiki/Scrum_(software_development). Accessed: 2023-01-11.

[9] Seed database. https://aurora.auburn.edu/handle/11200/49140. Accessed: 2023-07-12.

[10] Word error rate. https://en.wikipedia.org/wiki/Word_error_rate. Accessed: 2023-06-09.

[11] H. Albaqshi and A. Sagheer. Dysarthric speech recognition using convolutional recurrent neural networks. *International Journal of Intelligent Engineering & Systems*, 13(6), 2020.

[12] A. Amberkar, P. Awasarmol, G. Deshmukh, and P. Dave. Speech recognition using recurrent neural networks. In *2018 international conference on current trends towards converging technologies (ICCTCT)*, pages 1–4. IEEE, 2018.

[13] M. Bisani and H. Ney. Investigations on joint-multigram models for grapheme-to-phoneme conversion. In *INTERSPEECH*, 2002.

[14] M. Bisani and H. Ney. Joint-sequence models for grapheme-to-phoneme conversion. *Speech communication*, 50(5):434–451, 2008.

[15] B. Faghih, D. M. R. Azadehfar, P. Katebi, et al. User interface design for e-learning software. *arXiv preprint arXiv:1401.6365*, 2014.

[16] P. S. Ganney, S. Pisharody, and E. Claridge. Chapter 9 - software engineering. In A. Taktak, P. S. Ganney, D. Long, and R. G. Axell, editors, *Clinical Engineering (Second Edition)*, pages 131–168. Academic Press, second edition edition, 2020.

[17] J. S. Garofolo. Timit acoustic phonetic continuous speech corpus. *Linguistic Data Consortium, 1993*, 1993.

[18] L. Gelin, M. Daniel, J. Pinquier, and T. Pellegrini. End-to-end acoustic modelling for phone recognition of young readers. *Speech Communication*, 134:71–84, 2021.

[19] A. Graves, S. Fernández, and J. Schmidhuber. Bidirectional lstm networks for improved phoneme classification and recognition. In *Artificial Neural Networks: Formal Models and Their Applications–ICANN 2005: 15th International Conference, Warsaw, Poland, September 11-15, 2005. Proceedings, Part II 15*, pages 799–804. Springer, 2005.

[20] J. R. Green, R. L. MacDonald, P.-P. Jiang, J. Cattiau, R. Heywood, R. Cave, K. Seaver, M. A. Ladewig, J. Tobin, M. P. Brenner, P. C. Nelson, and K. Tomanek. Automatic Speech Recognition of Disordered Speech: Personalized Models Outperforming Human Listeners on Short Phrases. In *Proc. Interspeech 2021*, pages 4778–4782, 2021.

[21] F. Hamidi and M. Baljko. Automatic speech recognition: A shifted role in early speech intervention. *UMBC Information Systems Department Collection*, 2013.

[22] W. Han, C.-F. Chan, C.-S. Choy, and K.-P. Pun. An efficient mfcc extraction method in speech recognition. In *2006 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 4 pp.–, 2006.

[23] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[24] N. M. Joy and S. Umesh. Improving acoustic models in torgo dysarthric speech database. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(3):637–645, 2018.

[25] R. D. Kent, G. Weismer, J. F. Kent, and J. C. Rosenbek. Toward phonetic intelligibility testing in dysarthria. *Journal of Speech and Hearing Disorders*, 54(4):482–499, 1989.

[26] O. Khriyenko. Semantic ui: Automated creation of semantically personalized user interface. *GSTF Journal on Computing (JoC)*, 4:1–9, 2015.

[27] K. Kim, S. Trimi, H. Park, and S. Rhee. The impact of cms quality on the outcomes of e-learning systems in higher education: an empirical study. *Decision Sciences Journal of Innovative Education*, 10(4):575–587, 2012.

[28] D. Klakow and J. Peters. Testing the correlation of word error rate and perplexity. *Speech Communication*, 38(1-2):19–28, 2002.

[29] J. Kominek and A. W. Black. The cmu arctic speech databases. In *Fifth ISCA workshop on speech synthesis*, 2004.

[30] V. I. Levenshtein et al. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, volume 10, pages 707–710. Soviet Union, 1966.

[31] S. Li, M. S. Atkins, D. Bailey, J. Liu, Y. Cao, R. Bassy, and C. D. Seals. Software engineering to develop online phonetics educational training: interdisciplinary research with communications sciences and disorders. *American Society of Engineering Education (ASEE-SE 2020), Auburn AL*, page 11, 2020.

[32] T. Lin, Y. Wang, X. Liu, and X. Qiu. A survey of transformers. *AI Open*, 2022.

[33] J. Liu. Artificial intelligence-based enhancements supporting linguistic e-learning system. 2022.

[34] J. Liu, C. Ren, Y. Luan, S. Li, T. Xie, C. Seals, and M. Speights Atkins. Transformer-based multilingual g2p converter for e-learning system. In *Artificial Intelligence in HCI: 3rd International Conference, AI-HCI 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings*, pages 546–556. Springer, 2022.

[35] J. Liu, M. Speights, D. Bailey, S. Li, Y. Luan, I. Mishra, Y. Cao, and C. Seals. Optimization to automated phonetic transcription grading tool (aptgt)–automatic exam generator. In *Learning and Collaboration Technologies: New Challenges and Learning Experiences: 8th International Conference, LCT 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, Part I*, pages 80–91. Springer, 2021.

[36] B. MacWhinney. From childes to talkbank. 2001.

[37] B. MacWhinney. Talkbank: Building an open unified multimodal database of communicative interaction. 2004.

[38] B. MacWhinney. The talkbank project. In *Creating and Digitizing Language Corpora: Volume 1: Synchronic Databases*, pages 163–180. Springer, 2007.

[39] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[40] K. Rao, F. Peng, H. Sak, and F. Beaufays. Grapheme-to-phoneme conversion using long short-term memory recurrent neural networks. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4225–4229. IEEE, 2015.

[41] C. Ren, J. Liu, D. Feng, and C. D. Seals. Phone-based speech recognition for phonetic e-learning system. *HCI International Conference 2023, July 23–July 28, 2023, Proceedings*, 2023.

[42] F. Rudzicz, A. K. Namasivayam, and T. Wolff. The torgo database of acoustic and articulatory speech from speakers with dysarthria. *Language Resources and Evaluation*, 46:523–541, 2012.

[43] H. Sak, A. W. Senior, and F. Beaufays. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. 2014.

[44] C. D. Seals, S. Li, M. Speights Atkins, D. Bailey, J. Liu, Y. Cao, and R. Bassy. Applied webservices platform supported through modified edit distance algorithm: automated phonetic transcription grading tool (aptgt). In *Learning and Collaboration Technologies. Designing, Developing and Deploying Learning Experiences: 7th International Conference, LCT 2020, Held as Part of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings, Part I 22*, pages 380–398. Springer, 2020.

[45] L. D. Shriberg, J. Kwiatkowski, and K. Hoffmann. A procedure for phonetic transcription by consensus. *Journal of Speech, Language, and Hearing Research*, 27(3):456–465, 1984.

[46] A. Sokolov, T. Rohlin, and A. Rastrow. Neural machine translation for multilingual grapheme-to-phoneme conversion. *arXiv preprint arXiv:2006.14194*, 2020.

[47] M. Speights Atkins, D. J. Bailey, and S. E. Boyce. Speech exemplar and evaluation database (seed) for clinical training in articulatory phonetics and speech science. *Clinical Linguistics & Phonetics*, 34(9):878–886, 2020.

[48] M. Speights Atkins, D. J. Bailey, and C. D. Seals. Implementation of an automated grading tool for phonetic transcription training. *Clinical Linguistics & Phonetics*, pages 1–16, 2022.

[49] J. Titterington and S. Bates. Practice makes perfect? the pedagogic value of online independent phonetic transcription practice for speech and language therapy students. *Clinical Linguistics & Phonetics*, 32(3):249–266, 2018.

[50] W. M. Trochim and J. P. Donnelly. *Research methods knowledge base*, volume 2. Atomic Dog Pub. Macmillan Publishing Company, New York, 2001.

[51] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[52] R. Wiese. *The phonology of German.* Oxford University Press, USA, 2000.

[53] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, et al. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*, 2016.

[54] M. Xu, L.-Y. Duan, J. Cai, L.-T. Chia, C. Xu, and Q. Tian. Hmm-based audio keyword generation. In *Pacific-Rim Conference on Multimedia*, pages 566–574. Springer, 2004.

[55] S. Yang, Y. Wang, and X. Chu. A survey of deep learning techniques for neural machine translation. *arXiv preprint arXiv:2002.07526*, 2020.

[56] S. Yolchuyeva, G. Németh, and B. Gyires-Tóth. Transformer based grapheme-to-phoneme conversion. *arXiv preprint arXiv:2004.06338*, 2020.

[57] K. M. Yorkston, D. R. Beukelman, and C. Traynor. *Assessment of intelligibility of dysarthric speech*. Pro-ed Austin, TX, 1984.

Appendix A: Auburn IRB Approval Form

Memorandum

To: IRBadmin@auburn.edu

From: PI – Chang Ren czr0049@auburn.edu

CC: Co-I - Dr. Cheryl Seals sealscd@auburn.edu

Date: 06/07/2022

Subject: Revisions to IRB New Protocol

The exempt application and information letter have been carefully reviewed by the PI and Co-I and all revisions have been made in the attached documents following below:

- **1:** The current and start dates have been updated in the project identification.
- **4.f.** "The anticipated time it will take to use the applications is 20-40 minutes. The anticipated total time to complete all study activities is 30 – 60 minutes." These statements have been added to indicate the duration needed to use the applications and the total time needed to complete all study activities.
- "The estimated time required to complete task activities using both versions of the application ranges from 20 to 40 minutes." This statement has been added to the **Information Letter**. (Under the question of What will be involved if you participate?)
- **7.b.** "The participation in the study will not provide direct personal benefits to the participants." This statement has been added.
- "There are no personal benefits directly associated with participating in this research." This statement has been added to the **Information Letter**. (Under the question of Are there any benefits to yourself or others?)
- PI Chang has completed the Responsible Conduct of Research (RCR) training. The certificate of completion for the course has been included in the **Appendix of the CITI training documentation.**

All the requested changes have been done and highlighted.

Thank you.

Revised 10/18/2022

AUBURN UNIVERSITY HUMAN RESEARCH PROTECTION PROGRAM (HRPP)

EXEMPT REVIEW APPLICATION

For assistance, contact: **The Office of Research Compliance (ORC)**
Phone: **334-844-5966** E-Mail: **IRBAdmin@auburn.edu** Web Address: **http://www.auburn.edu/research/vpr/ohs**
Submit completed form and supporting materials as one PDF through the IRB Submission Page
Hand written forms are not accepted. Where links are found hold down the control button (Ctrl) then click the link..

1. Project Identification **Today's Date: June 7, 2023**

Anticipated start date of the project: June 15, 2023 **Anticipated duration of project: 1 Year**

a. **Project Title: Enhancing User Experience through improving the User Interface of phonetics tools and studies on phone-level ASR-based automation through deep learning techniques**

b. **Principal Investigator (PI): Chang Ren** Degree(s): Click or tap here to enter text.
Rank/Title: Graduate Student Department/School: Computer Science and Software Engineering
Role/responsibilities in this project: **PI**
Preferred Phone Number: **3342755077** AU Email: czr0049@auburn.edu

Faculty Advisor Principal Investigator (if applicable): Cheryl Seals
Rank/Title: Professor Department/School: Computer Science and Software Engineering
Role/responsibilities in this project: **Co-I**
Preferred Phone Number: **3348446319** AU Email: sealscd@auburn.edu

Department Head: Hari Narayanan Department/School: Computer Science and Software Engineering
Preferred Phone Number: **3348446312** AU Email: naraynh@auburn.edu
Role/responsibilities in this project: Click or tap here to enter text.

c. **Project Key Personnel –** Identify all key personnel who will be involved with the conduct of the research and describe their role in the project. Role may include design, recruitment, consent process, data collection, data analysis, and reporting. (To determine key personnel, see decision tree**).** *Exempt determinations are made by individual institutions; reliance on other institutions for exempt determination is not feasible. Non-AU personnel conducting exempt research activities must obtain approval from the IRB at their home institution.*

Key personnel are required to maintain human subjects training through CITI. Only for EXEMPT level research is documentation of completed CITI training NO LONGER REQUIRED to be included in the submission packet. NOTE however, **the IRB will perform random audits of CITI training records to confirm** reported training courses and expiration dates. Course title and expiration dates are shown on training certificates.

Name: Chang Ren Degree(s): Click or tap here to enter text.
Rank/Title: Graduate Student Department/School: Computer Science and Software Engineering
Role/responsibilities in this project: PI Ren will be responsible for conducting research, designing and implementing the necessary applications, developing protocols and surveys, carrying out experiments, recruiting and interacting with participants.
- AU affiliated? ☒ Yes ☐ No If no, name of home institution: Click or tap here to enter text.
- Plan for IRB approval for non-AU affiliated personnel? Click or tap here to enter text.
- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No
- If yes, briefly describe the potential or real conflict of interest: Click or tap here to enter text.
- Completed required CITI training? ☒ Yes ☐ No If NO, complete the appropriate CITI basic course and update the revised Exempt Application form.

- If YES, choose course(s) the researcher has completed: Conflicts of Interest in Research Involving Human Subjects(2024), Defining Research with Human Subjects(2024), History and Ethical Principles(2024), IRB # 2 Social and Behavioral Emphasis(2024), Research in Public Elementary and Secondary Schools (2025), Internet Research(2026), AU Basic RCR Training for ALL Faculty, Staff, Postdocs, and Students (2026).

Name: Dr. Cheryl Seals Degree(s): Click or tap here to enter text.
Rank/Title: Professor Department/School: Computer Science and Software Engineering
Role/responsibilities in this project: Co-I Seals will assist with research prepare protocols, surveys and assist experimental trials and meet with participants.
- AU affiliated? ☒ Yes ☐ No If no, name of home institution: Click or tap here to enter text.
- Plan for IRB approval for non-AU affiliated personnel? Click or tap here to enter text.
- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No
- If yes, briefly describe the potential or real conflict of interest: Click or tap here to enter text.
- Completed required CITI training? ☐ Yes ☐ No If NO, complete the appropriate CITI basic course and update the revised EXEMPT application form.
- If YES, choose course(s) the researcher has completed: Conflicts of Interest in Research Involving Human Subjects (2024), IRB # 2 Social and Behavioral Emphasis (2025), History and Ethical Principles (2024), Responsible Conduct of Research for Social and Behavioral (2027).

Name: Click or tap here to enter text. Degree(s): Click or tap here to enter text.
Rank/Title: Choose Rank/Title Department/School: Choose Department/School
Role/responsibilities in this project: Click or tap here to enter text.
- AU affiliated? ☐ Yes ☐ No If no, name of home institution: Click or tap here to enter text.
- Plan for IRB approval for non-AU affiliated personnel? Click or tap here to enter text.
- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☐ No
- If yes, briefly describe the potential or real conflict of interest: Click or tap here to enter text.
- Completed required CITI training? ☐ Yes ☐ No If NO, complete the appropriate CITI basic course and update the revised EXEMPT application form.
- If YES, choose course(s) the researcher has completed: Choose a course Expiration Date
Choose a course Expiration Date

d. Funding Source – Is this project funded by the investigator(s)? Yes ☒ No ☐
Is this project funded by AU? Yes ☐ No ☒ If YES, identify source Click or tap here to enter text.
Is this project funded by an external sponsor? Yes ☐ No ☒ If YES, provide name of sponsor, type of sponsor (governmental, non-profit, corporate, other), and an identification number for the award.
Name: Click or tap here to enter text. Type: Click or tap here to enter text. Grant #: Click or tap here to enter text.

e. List other AU IRB-approved research projects and/or IRB approvals from other institutions that are associated with this project. Describe the association between this project and the listed project(s):
Click or tap here to enter text.

2. Project Summary

a. Does the study *TARGET* any special populations? Answer YES or NO to all.

Minors (under 18 years of age; if minor participants, at least 2 adults must be present during all research procedures that include the minors)	Yes ☐ No ☒
Auburn University Students	Yes ☒ No ☐
Pregnant women, fetuses, or any products of conception	Yes ☐ No ☒

Prisoners or wards (unless incidental, not allowed for Exempt research) Yes ☐ No ☒

Temporarily or permanently impaired Yes ☐ No ☒

b. Does the research pose more than minimal risk to participants? Yes ☐ No ☒

If YES, to question 2.b, then the research activity is NOT eligible for EXEMPT review. Minimal risk means that the probability and magnitude of harm or discomfort anticipated in the research is not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or test. 42 CFR 46.102(i)

c. Does the study involve any of the following? *If YES to any of the questions in item 2.c, then the research activity is NOT eligible for EXEMPT review.*

Procedures subject to FDA regulations (drugs, devices, etc.) Yes ☐ No ☒

Use of school records of identifiable students or information from instructors about specific students. Yes ☐ No ☒

Protected health or medical information when there is a direct or indirect link which could identify the participant. Yes ☐ No ☒

Collection of sensitive aspects of the participant's own behavior, such as illegal conduct, drug use, sexual behavior or alcohol use. Yes ☐ No ☒

d. Does the study include deception? Requires limited review by the IRB* Yes ☐ No ☒

3. MARK the category or categories below that describe the proposed research. Note the IRB Reviewer will make the final determination of the eligible category or categories.

☒ **1.** Research conducted in established or commonly accepted educational settings, involving normal educational practices. The research is not likely to adversely impact students' opportunity to learn or assessment of educators providing instruction. 104(d)(1)

☒ **2.** Research only includes interactions involving educational tests, surveys, interviews, public observation if at least ONE of the following criteria. (The research includes data collection only; may include visual or auditory recording; may NOT include intervention and only includes interactions). **Mark the applicable sub-category below (I, ii, or iii). 104(d)(2)**

☒ **(i)** Recorded information cannot readily identify the participant (directly or indirectly/ linked);
OR
- surveys and interviews: no children;
- educational tests or observation of public behavior: can only include children when investigators do not participate in activities being observed.

☒ **(ii)** Any disclosures of responses outside would not reasonably place participant at risk; **OR**

☐ **(iii)** Information is recorded with identifiers or code linked to identifiers and IRB conducts limited review; no children. **Requires limited review by the IRB.***

☐ **3.** Research involving Benign Behavioral Interventions (BBI)** through verbal, written responses including data entry or audiovisual recording from adult subjects who prospectively agree and ONE of the following criteria is met. (This research does not include children and does not include medical interventions. Research cannot have deception unless the participant prospectively agrees that they will be unaware of or misled regarding the nature and purpose of the research) **Mark the applicable sub-category below (A, B, or C).** 104(d)(3)(i)

☐ **(A)** Recorded information cannot readily identify the subject (directly or indirectly/ linked); **OR**

☐ **(B)** Any disclosure of responses outside of the research would not reasonably place subject at risk; **OR**

☐ **(C)** Information is recorded with identifies and cannot have deception unless participants prospectively agree. **Requires limited review by the IRB.***

☐ **4.** Secondary research for which consent is not required: use of identifiable information or identifiable bio-specimen that have been or will be collected for some other 'primary' or 'initial' activity, if one of the following criteria is met. Allows retrospective and prospective secondary use. **Mark the applicable sub-category below (i, ii, iii, or iv).** 104 (d)(4)

☐ **(i)** Bio-specimens or information are publicly available;

☐ **(ii)** Information recorded so subject cannot readily be identified, directly or indirectly/linked investigator does not contact subjects and will not re-identify the subjects; **OR**

☐ **(iii)** Collection and analysis involving investigators use of identifiable health information when us is regulated by HIPAA "health care operations" or "research" or "public health activities and purposes" (does not include bio-specimens (only PHI and requires federal guidance on how to apply); **OR**

☐ **(iv)** Research information collected by or on behalf of federal government using government generated or collected information obtained for non-research activities.

☐ **5.** Research and demonstration projects which are supported by a federal agency/department AND designed to study and which are designed to study, evaluate, or otherwise examine: (i)public benefit or service programs; (ii) procedures for obtaining benefits or services under those programs; (iii) possible changes in or alternatives to those programs or procedures; or (iv) possible changes in methods or levels of payment for benefits or service under those programs. (must be posted on a federal web site). 104.5(d)(5) (must be posted on a federal web site)

☐ **6.** Taste and food quality evaluation and consumer acceptance studies, (i) if wholesome foods without additives and consumed or (ii) if a food is consumed that contains a food ingredient at or below the level and for a use found to be safe, or agricultural chemical or environmental contaminant at or below the level found to be safe, by the Food and Drug Administration or approved by the Environmental Protection Agency or the Food Safety and Inspection Service of the U.S. Department of Agriculture. The research does not involve prisoners as participants. 104(d)(6)

**Limited IRB review – the IRB Chair or designated IRB reviewer reviews the protocol to ensure adequate provisions are in place to protect privacy and confidentiality.*

***Category 3 – Benign Behavioral Interventions (BBI) must be brief in duration, painless/harmless, not physically invasive, not likely to have a significant adverse lasting impact on participants, and it is unlikely participants will find the interventions offensive or embarrassing.*

**** Exemption categories 7 and 8 require broad consent. The AU IRB has determined the regulatory requirements for legally effective broad consent are not feasible within the current institutional infrastructure. EXEMPT categories 7 and 8 will not be implemented at this time.*

4. Describe the proposed research including who does what, when, where, how, and for how long, etc.

a. Purpose

The purpose of this study is to enhance the user interface design and user experience of the Automated Phonetic Transcription Grading Tool (APTgt), an interactive web-based E-learning system designed to support communication disorders faculty in teaching phonetic transcription. The research aims to optimize the current designs by improving ease of use, aesthetics, and consistency in design, with the ultimate goal of providing a more efficient and user-friendly tool for linguistic instructors. The study specifically addresses the following questions: How can the user interface of APTgt be optimized to improve ease of use, aesthetics, and consistency in design? How can the efficiency of APTgt be increased to reduce the time required for instructors to use the software?

b. Participant population, including the number of participants and the rationale for determining number of participants to recruit and enroll. Note if the study enrolls minor participants, describe the process to ensure more than 1 adult is present during all research procedures which include the minor.

The participants will be linguistic professionals (i.e. advanced undergraduate, graduates or faculty in a field that is practiced in linguistic transcription) and usability experts who possess usability certification, have completed a course in usability, or have relevant experience as a usability researcher or member of technical staff). We will employ email and opportunistic sampling as well as snowball recruiting methods to determine the number of participants we would like to recruit.

c. Recruitment process. Address whether recruitment includes communications/interactions between study staff and potential participants either in person or online. *Submit a copy of all recruitment materials.*

We will recruit using email to linguistic instructors that we have worked with on prior projects, opportunistic sampling, and snowball recruiting. We will provide an example recruitment email in the IRB appendix and will send them a survey link: https://auburn.qualtrics.com/jfe/form/SV_2i3EltzcU7HCn5k

d. Consent process including how information is presented to participants, etc.

The information letter will be used for consent.

e. Research procedures and methodology

To evaluate the effectiveness of the enhancements, participants will be provided with two versions of the APTgt application—an existing benchmark version and an experimental version with the proposed improvements. The study will involve conducting a comparative usability evaluation. Participants will first complete a pre-questionnaire to gather their initial perceptions and expectations of the application (5-10 minutes). They will then use both versions of the application for a duration of 20-40 minutes. Finally, the session will conclude with a post-questionnaire to gather feedback on aesthetics, ease of use, satisfaction, and overall usability of the application (5-10 minutes).

f. Anticipated time per study exercise/activity and total time if participants complete all study activities.

Duration of time needed to complete the pre-survey is 5 - 10 minutes and post-survey is 5 - 10 minutes. The anticipated time it will take to use the applications is 20-40 minutes. The anticipated total time to complete all study activities is 30 – 60 minutes.

g. Location of the research activities.

The research will be done online.

h. Costs to and compensation for participants? If participants will be compensated describe the amount, type, and process to distribute.

There will be no compensations.

i. Non-AU locations, site, institutions. *Submit a copy of agreements/IRB approvals.*
Click or tap here to enter text.

j. Describe how results of this study will be used (presentation? publication? thesis? dissertation?)
The result of this study will be used for a Ph.D. dissertation.

k. Additional relevant information.
Click or tap here to enter text.

5. Waivers
Check applicable waivers and describe how the project meets the criteria for the waiver.

☐ Waiver of Consent (Including existing de-identified data)

☒ Waiver of Documentation of Consent (Use of Information Letter, rather than consent form requiring signatures)

☐ Waiver of Parental Permission (in Alabama, 18 years-olds may be considered adults for research purposes)
https://sites.auburn.edu/admin/orc/irb/IRB 1 Exempt and Expedited/11-113 MR 1104 Hinton Renewal 2021-1.pdf

a. Provide the rationale for the waiver request.
We have updated the Informed Consent to an Information Letter by eliminating unnecessary signatures. The data collected will be anonymous for those who complete pre-survey only. For those who choose to continue with the experience, we will assign a code that links their pre-to-post survey results. (i.e. the data will be collected and stored on Qualtrics servers and Auburn Box to ensure confidentiality. Access to the survey results will be limited to the Primary Investigator and Co-Investigator only.).

6. Describe the process to select participants/data/specimens. If applicable, include gender, race, and ethnicity of the participant population.

The participants will be linguistic professionals and usability experts who are undergraduate, graduate students or faculty at Auburn University and are above 18 years old. We will recruit participants of all genders and distribute the surveys online via Qualtrics.

7. Risks and Benefits

7a. Risks - Describe why none of the research procedures would cause a participant either physical or psychological discomfort or be perceived as discomfort above and beyond what the person would experience in daily life (minimal risk).

There are no risks associated with this research.

7b. Benefits – Describe whether participants will benefit directly from participating in the study. If yes, describe the benefit. And, describe generalizable benefits resulting from the study.

The participation in the study will not provide direct personal benefits to the participants. However, participating students may gain advanced knowledge in the area of User Interface Design, while faculty members may benefit from enhanced teaching experiences using the E-learning system.

8. Describe the provisions to maintain confidentiality of data, including collection, transmission, and storage. Identify platforms used to collect and store study data. *For EXEMPT research, the AU IRB recommends AU BOX or using an AU issued and encrypted device. If a data collection form will be used, submit a copy.*

The survey data will be securely stored on Qualtrics servers and will be stored on Auburn Box. Only Primary Investigator and Co-Investigators have access to the survey results.

a. If applicable, submit a copy of the data management plan or data use agreement.

9. Describe the provisions included in the research to protect the privacy interests of participants (e.g., others will not overhear conversations with potential participants, individuals will not be publicly identified or embarrassed).

Participant data will be collected anonymously, with no connection to identifying information. The survey data will be securely stored on Qualtrics servers and Auburn Box, and only the Primary Investigator and Co-Investigators will have access to the survey results.

10. Does this research include purchase(s) that involve technology hardware, software or online services?

☐ **YES** ☒ **NO**

If YES:

A. **Provide the name of the product** Click or tap here to enter text.
and the manufacturer of the product Click or tap here to enter text.
B. **Briefly describe use of the product in the proposed human subject's research.**
Click or tap here to enter text.
C. **To ensure compliance with AU's Electronic and Information Technology Accessibility Policy, contact AU IT Vendor Vetting team at vetting@auburn.edu to learn the vendor registration process (prior to completing the purchase).**
D. **Include a copy of the documentation of the approval from AU Vetting with the revised submission.**

11. Additional Information and/or attachments.

In the space below, provide any additional information you believe may help the IRB review of the proposed research. If attachments are included, list the attachments below. Attachments may include recruitment materials, consent documents, site permissions, IRB approvals from other institutions, data use agreements, data collection form, CITI training documentation, etc.

Information Letter
Recruitment Email
Printed version of online survey. (https://auburn.qualtrics.com/jfe/form/SV_2i3EltzcU7HCn5k)
Chang Ren Citi Training
Cheryl Seals Citi Training

Required Signatures *(If a student PI is identified in item 1.a, the EXEMPT application must be re-signed and updated at every revision by the student PI and faculty advisor. The signature of the department head is required only on the initial submission of the EXEMPT application, regardless of PI. Staff and faculty PI submissions require the PI signature on all* *version, the department head signature on the original submission)*

Signature of Principal Investigator: Chang Ren **Date:** 06/07/2023

Signature of Faculty Advisor (If applicable): [signature] **Date:** 06/08/2023

Signature of Dept. Head: N. Hui Nguyen **Date:** 5/26/23

Version Date: **6/7/2023**

AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE AND SOFTWARE ENGINEERING

(NOTE: DO NOT AGREE TO PARTICIPATE UNLESS IRB APPROVAL INFORMATION WITH CURRENT DATES HAS BEEN ADDED TO THIS DOCUMENT.)

INFORMATION LETTER
for a Research Study entitled
"Enhancing User Experience through improving the User Interface of phonetics tools and studies on phone-level ASR-based automation through deep learning techniques"

You are invited to participate in a research study to enhance the user interface and user experience of the Automated Phonetic Transcription Grading Tool (APTgt), an interactive web-based E-learning system designed to support communication disorders faculty in teaching phonetic transcription. The goal is to optimize the current designs by improving ease of use, aesthetics, and design consistency. The study is being conducted by Chang Ren, PhD candidate, under the direction of Cheryl Seals, Charles W. Barkley Professor in the Auburn University Department of Computer Science and Software Engineering. You are invited to participate because you are Linguistic professionals or Usability experts and are age 19 or older.

What will be involved if you participate?
Your participation is completely voluntary. If you decide to participate in this research study, you will be asked to complete an online survey about your educational background, qualifications, specific areas of expertise, and previous experience with E-learning systems or user interface design. Your total time commitment will be approximately 5 – 10 minutes.
To continue participating in the follow-up experiment and task activities, you can express your interest by providing your contact email on a separate form that does not collect any identifying information. After completing the initial survey, you will be directed to an online form where you can enter your email address. If you do not wish to be contacted, you can simply close the tab containing the form. Participants who indicate their willingness to be selected will be contacted for task activities and an online post-questionnaire once we are ready to gather feedback on application evaluation. During this process, participants will engage in usability evaluation of the application and complete a post-questionnaire. The estimated time required to complete task activities using both versions of the application is anticipated to be between 20 to 40 minutes. Post- Questionnaire take approximately 5–10 minutes.

Are there any risks or discomforts? There are no risks or discomforts involved.

Are there any benefits to yourself or others?
There are no personal benefits directly associated with participating in this research. However, if you participate in this study, you can expect to contribute your expertise to enhance the user interface and user experience of phonetic E-learning system. The involvement will directly influence the system's improvement, making it more efficient and user-friendly for linguistic instructors. Additionally, participants will gain hands-on experience in usability evaluation, enhancing your professional skills and expertise in user-centered design. Your contributions will shape the future of E-learning in communication disorders, benefiting educators and students in the field.

Will you receive compensation for participating? There is no compensation for participation.
Are there any costs? There are no costs to participate in the research.

If you change your mind about participating, you can withdraw at any time by stopping the survey. Then the data will not be collected or recorded.

Any data obtained in connection with this study will remain anonymous and confidential. The study will take place via an online survey. Collected data with Auburn Qualtrics will be stored on Auburn Box. Only the investigators of this research will have access to the data obtained.

If you have questions about this study, please contact Chang Ren at czr0049@auburn.edu or Dr. Cheryl Seals at sealscd@auburn.edu.

If you have questions about your rights as a research participant, you may contact the Auburn University Office of Research Compliance or the Institutional Review Board by phone (334) 844-5966 or e-mail at IRBadmin@auburn.edu or IRBChair@auburn.edu.

HAVING READ THE INFORMATION ABOVE, YOU MUST DECIDE IF YOU WANT TO PARTICIPATE IN THIS RESEARCH PROJECT. IF YOU DECIDE TO PARTICIPATE, PLEASE CLICK ON THE LINK BELOW.
YOU MAY PRINT A COPY OF THIS LETTER TO KEEP.

Chang Ren ______________ 06/07/2023
Investigator Date

Chang Ren
Printed Name

Cheryl P. Seals ______________ 06/08/2023
Co-Investigator Date

Dr. Cheryl Seals
Printed Name

LINK TO SURVEY

Version Date (date document created): 06/07/2023

RECRUITMENT EMAIL

Dear potential participants,

Thank you for taking the time to read this email. My name is Chang Ren, a Ph.D. candidate studying Computer Science and Software Engineering at Auburn University. Under the guidance of my advisor, Dr. Cheryl Seals, Professor in the Auburn University Department of Computer Science and Software Engineering, we are conducting a survey targeting linguistic professionals and usability experts to evaluate the optimized user interface design and user experience of the Automated Phonetic Transcription Grading Tool (APTgt), an online E-learning system designed to support communication disorders faculty.

Your participation is completely voluntary. If you decide to participate in this research study, you will be asked to complete an online survey about your educational background, qualifications, specific areas of expertise, and previous experience with E-learning systems or user interface design. Then we are specifically interested in gathering opinion and feedback on ease of use, aesthetics, consistency in design, and overall usability of the application. Your insights will be instrumental in optimizing the user interface design to provide a more efficient and user-centered experience for linguistic instructors. Your commitment time for this survey will be approximately 5 - 10 minutes.

If you have questions about this study, please contact Chang Ren at czr0049@auburn.edu or Dr. Cheryl Seals at sealscd@auburn.edu. You can find the full Information Letter attached below.

SURVEY LINK

https://auburn.qualtrics.com/jfe/form/SV_2i3EltzcU7HCn5k

Thank you again for considering participation in our research. Your input and expertise are greatly appreciated.

Best regards,
Chang Ren
Ph.D. Candidate
Department of Computer Science and Software Engineering
Auburn University, Auburn, AL

PRE-SURVEY

Survey Completion

0% 100%

AUBURN UNIVERSITY

Hello! **You are invited to participate in a research study** to enhance the user interface and user experience of the Automated Phonetic Transcription Grading Tool (APTgt), an interactive web-based E-learning system designed to support communication disorders faculty in teaching phonetic transcription. The goal is to optimize the current designs by improving ease of use, aesthetics, and design consistency. The study is being conducted by Chang Ren, PhD candidate, under the direction of Cheryl Seals, Charles W. Barkley Professor in the Auburn University Department of Computer Science and Software Engineering. You are invited to participate because you are Linguistic professionals or Usability experts and are age 19 or older.

What will be involved if you participate? Your participation is completely voluntary. If you decide to participate in this research study, you will be asked to complete an online survey about your educational background, qualifications, specific areas of expertise, and previous experience with E-learning systems or user interface design. Your total time commitment will be approximately 5 – 10 minutes. To continue participating in the follow-up experiment and task activities, you can express your interest by providing your contact email on a separate form that does not collect any identifying information. After completing the initial survey, you will be directed to an online form where you can enter your email address. If you do not wish to be contacted, you can close the tab containing the form. Participants who indicate their willingness to be selected will be contacted for task activities and an online post-questionnaire once we are ready to gather feedback on the application evaluation. During this process, participants will engage in a usability evaluation of the application and complete a post-questionnaire. Post-Questionnaire takes approximately 5–10 minutes.

Are there any risks or discomforts? There are no risks or discomforts involved.

Are there any benefits to yourself or others? If you participate in this study, you can expect to contribute your expertise to enhance the user interface and user experience of the phonetic E-learning system. The involvement will directly influence the system's improvement, making it more efficient and user-friendly for linguistic instructors. Additionally, participants will gain hands-on experience in usability evaluation, enhancing their professional skills and expertise in user-centered design. Your contributions will shape the future of E-learning in communication disorders, benefiting educators and students in the field.

Will you receive compensation for participating? There is no compensation for participation.

Are there any costs? There are no costs to participate in the research.

If you change your mind about participating, you can withdraw at any time by stopping the survey. Then the data will not be collected or recorded.

Any data obtained in connection with this study will remain anonymous and confidential. The study will take place via an online survey. Collected data with Auburn Qualtrics will be stored on Auburn Box. Only the investigators of this research will have access to the data obtained.

If you have questions about this study, please contact Chang Ren at czr0049@auburn.edu or Dr. Cheryl Seals at sealscd@auburn.edu.

If you have questions about your rights as a research participant, you may contact the Auburn University Office of Research Compliance or the Institutional Review Board by phone (334) 844-5966 or e-mail at IRBadmin@auburn.edu or IRBChair@auburn.edu.

HAVING READ THE INFORMATION ABOVE, YOU MUST DECIDE IF YOU WANT TO PARTICIPATE IN THIS RESEARCH PROJECT. IF YOU DECIDE TO PARTICIPATE, PLEASE CLICK ON THE LINK BELOW. YOU MAY PRINT A COPY OF THIS LETTER TO KEEP.

››

Powered by Qualtrics

AUBURN UNIVERSITY

What is your gender?

- ◯ Female
- ◯ Male
- ◯ Not Listed

How old are you?

- ◯ 19 - 24
- ◯ 25 - 34
- ◯ 35 +

What is your current role?

- ◯ Undergradute Student
- ◯ Graduate Student
- ◯ Professional

What are your areas of expertise? *Select all that apply.*

- ☐ Linguistics
- ☐ Usability
- ☐ Other (specify)

How many years of experience do you have in your field?

- ◯ 0 - 2 years
- ◯ 3 - 5 years
- ◯ 6 - 10 years
- ◯ 10 + years

Do you have prior experience working in an E-learning system?

- ○ Yes
- ○ No

How familiar you are with E-learning system?

- ○ Very familiar, I have extensive experience using E-learning systems.
- ○ Moderately familiar, I have some experience using E-learning systems.
- ○ Somewhat familiar, I have limited experience using E-learning systems.

what is your most frequently used E-learning system? *Select all that apply.*

- ☐ Canvas
- ☐ Moodle
- ☐ Blackboard Learn
- ☐ Coursera
- ☐ Udemy
- ☐ Other (specify)

Are there any specific challenges you have encountered when using E-learning systems? *Select all that apply.*

- ☐ Unappealing
- ☐ Inconsistent design
- ☐ Cluttered/complex interface
- ☐ Poor use of colors/typography/layout.
- ☐ Insufficient user control
- ☐ Navigation difficulties
- ☐ Lack of feedback
- ☐ Other (specify)

Do you have prior experience working in an E-learning system?

- ◯ Yes
- ◯ No

What are your goals for using APTgt? *Select all that apply.*

- ☐ Manage online courses/exams in linguistics.
- ☐ Improve my skills in phonetic transcription.
- ☐ Conduct research in communication disorders.
- ☐ Other (specify)

How often do you access APTgt?

- ◯ Daily
- ◯ A few times per week
- ◯ A few times per month
- ◯ Once or twice a semester or less often

Are there any specific challenges you have encountered when using APTgt? *Select all that apply.*

- ☐ Unappealing
- ☐ Inconsistent design
- ☐ Cluttered/complex interface
- ☐ Poor use of colors/typography/layout.
- ☐ Insufficient user control
- ☐ Navigation difficulties
- ☐ Lack of feedback
- ☐ Other (specify)

How likely are you to use APTgt again in the future?

- ○ Extremely unlikely
- ○ Somewhat unlikely
- ○ Neither likely nor unlikely
- ○ Somewhat likely
- ○ Extremely likely

<< >>

Have you been involved in usability evaluations in the past?

- ◉ Yes
- ○ No

How many usability evaluations have you been involved in?

- ○ 1-5
- ○ 6-10
- ○ 11-15
- ○ 16-20
- ○ 21+

Which usability evaluation methods are you most experienced in? *Select all that apply.*

- ☐ User interviews
- ☐ User testing
- ☐ Heuristic evaluations
- ☐ Cognitive walkthroughs
- ☐ Surveys/questionnaires
- ☐ Other (specify)

In which domains have you primarily experienced in usability evaluations? *Select all that apply.*

- ☐ Software
- ☐ Web application
- ☐ Mobile
- ☐ Education
- ☐ Healthcare
- ☐ E-commerce
- ☐ Gaming
- ☐ Manufacturing
- ☐ Retail
- ☐ Other (please specify)

Have you involved in usability evaluations for E-learning systems in the past?

- ○ Yes, multiple times
- ○ Yes, once or twice
- ○ No.

Thank you so much for your time spent taking this survey.

If you would like to be contacted in the future for feedback or evaluation of the system, please proceed to the end of the survey where you will be redirected to an online form to enter your email address.

If you do not wish to be contacted, please continue to the end of the survey. After you are redirected, please simply close the webpage to opt out.

<< >>

CONTACT FORM

Please provide your email address for follow-up.

1. Email

Submit

Create Your Own Form

forms.app

POST SURVEY

AUBURN
UNIVERSITY

Please select your role or select "Other."

- ○ Linguistic Professional
- ○ Usability Expert
- ○ Other (please specify)

Overall, what's your experience been with the APTgt v2.0?

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
The application is more visually pleasing than old version.	○	○	○	○	○
The application is easier to use than old version.	○	○	○	○	○
The navigation is more efficient than old version.	○	○	○	○	○
I am more satisfied with the application than the old version.	○	○	○	○	○

In the scale of 1-5, please rate the APTgt v2.0 in terms of the following aspects?

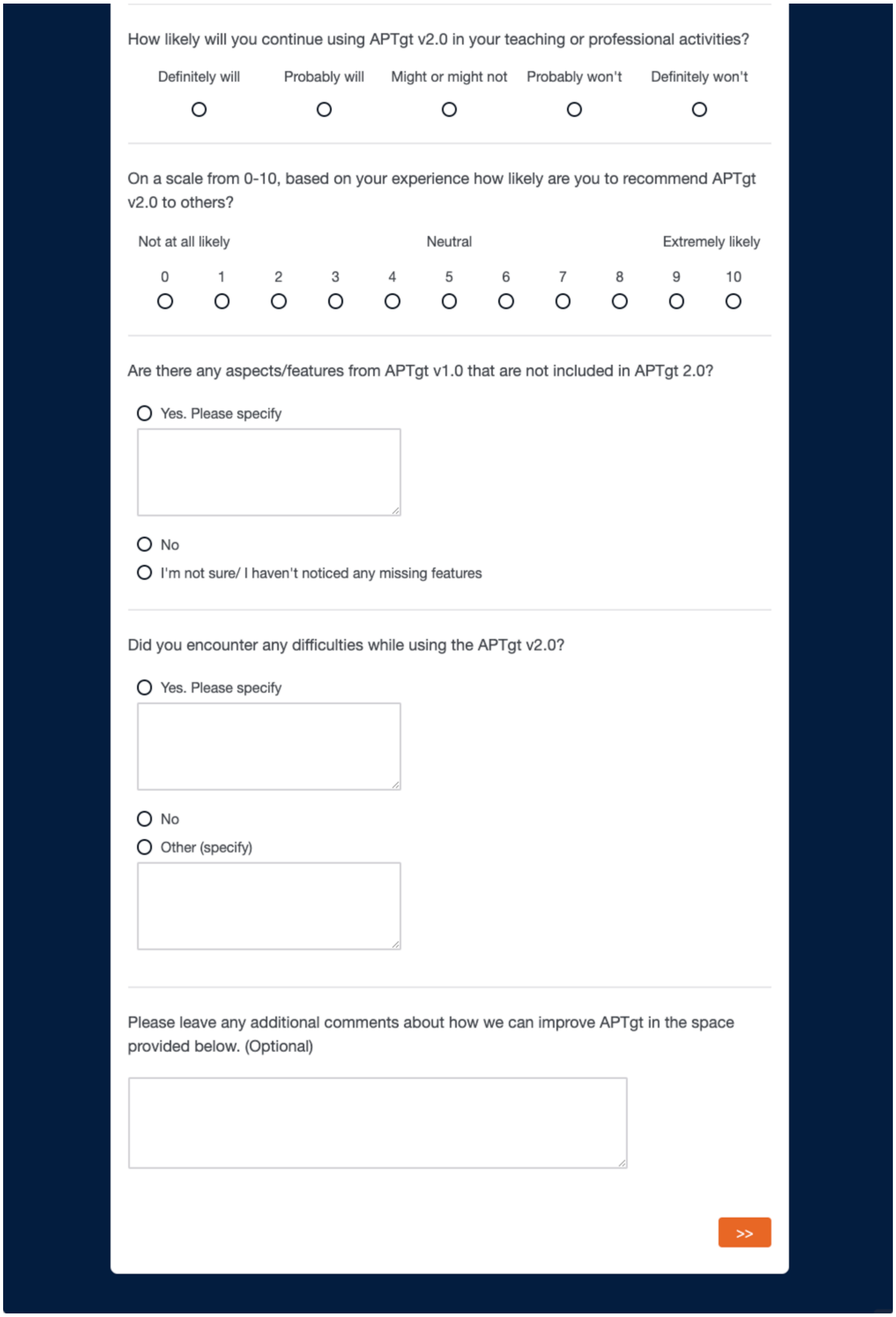
	Excellent	Very Good	Good	Poor	Very poor
Layout	○	○	○	○	○
Use of color	○	○	○	○	○
Consistency	○	○	○	○	○
Easy to use	○	○	○	○	○
Navigation	○	○	○	○	○

The APTgt v2.0 help to complete tasks more efficiently than the old version?

Strongly Agree	Somewhat agree	Neutral	Somewhat disagree	Strongly disagree
○	○	○	○	○

How would you rate the overall usability of APTgt v2.0 compared to the old version?

Much more usable	Somewhat more usable	Same	Somewhat less usable	Much less usable
○	○	○	○	○


How likely will you continue using APTgt v2.0 in your teaching or professional activities?

Definitely will	Probably will	Might or might not	Probably won't	Definitely won't
○	○	○	○	○

On a scale from 0-10, based on your experience how likely are you to recommend APTgt v2.0 to others?

Not at all likely — Neutral — Extremely likely

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Are there any aspects/features from APTgt v1.0 that are not included in APTgt 2.0?

- ○ Yes. Please specify
- ○ No
- ○ I'm not sure/ I haven't noticed any missing features

Did you encounter any difficulties while using the APTgt v2.0?

- ○ Yes. Please specify
- ○ No
- ○ Other (specify)

Please leave any additional comments about how we can improve APTgt in the space provided below. (Optional)

>>

The following survey questions will be added to the post survey for more comprehensive insights from evaluators.

1. Overall, I am satisfied with the ease of completing the tasks in this scenario. (Strongly disagree – Strongly agree)
2. Overall, I am satisfied with the amount of time it took to complete the tasks in this scenario. (Strongly disagree – Strongly agree)
3. It's user-friendly. (Strongly disagree – Strongly agree)
4. The interface of this system is pleasant
5. I can use it without written instructions. (Strongly disagree – Strongly agree)
6. It is pleasant to use. (Strongly disagree – Strongly agree)
7. It works the way I want it to work. (Strongly disagree – Strongly agree)
8. My interaction with the system would be clear and understandable (unlikely – likely)

CITI PROGRAM

Completion Date 24-Feb-2021
Expiration Date 24-Feb-2024
Record ID 25863904

This is to certify that:

Chang Ren

Not valid for renewal of certification through CME.

Has completed the following CITI Program course:

IRB Additional Modules
(Curriculum Group)
Conflicts of Interest in Research Involving Human Subjects
(Course Learner Group)
1 - Basic Course
(Stage)

Under requirements set by:

Auburn University

CITI
Collaborative Institutional Training Initiative

101 NE 3rd Avenue, Suite 320
Fort Lauderdale, FL 33301 US
www.citiprogram.org

Verify at www.citiprogram.org/verify/?w4fe0b188-9ac4-4763-9f29-24a9ce366e50-25863904

Completion Date 24-Feb-2021
Expiration Date 24-Feb-2024
Record ID 25863906

This is to certify that:

Chang Ren

Has completed the following CITI Program course:

Not valid for renewal of certification through CME.

IRB Additional Modules
(Curriculum Group)
Defining Research with Human Subjects - SBE
(Course Learner Group)
1 - Basic Course
(Stage)

Under requirements set by:

Auburn University

CITI
Collaborative Institutional Training Initiative

101 NE 3rd Avenue, Suite 320
Fort Lauderdale, FL 33301 US
www.citiprogram.org

Verify at www.citiprogram.org/verify/?w41755bce-1025-4028-800a-a51df2c4d5eb-25863906

CITI PROGRAM

Completion Date 24-Feb-2021
Expiration Date 24-Feb-2024
Record ID 39055407

This is to certify that:

Chang Ren

Has completed the following CITI Program course:

Not valid for renewal of certification through CME.

IRB Additional Modules
(Curriculum Group)
History and Ethical Principles - SBE
(Course Learner Group)
1 - Basic Course
(Stage)

Under requirements set by:

Auburn University

CITI
Collaborative Institutional Training Initiative

101 NE 3rd Avenue, Suite 320
Fort Lauderdale, FL 33301 US
www.citiprogram.org

Verify at www.citiprogram.org/verify/?w8238f834-ed39-4849-8a98-31c1bf2c2eca-39055407

Completion Date 24-Feb-2021
Expiration Date 24-Feb-2024
Record ID 25863903

This is to certify that:

Chang Ren

Has completed the following CITI Program course:

Not valid for renewal of certification through CME.

IRB # 2 Social and Behavioral Emphasis - AU Personnel - Basic/Refresher
(Curriculum Group)
IRB # 2 Social and Behavioral Emphasis - AU Personnel
(Course Learner Group)
1 - Basic Course
(Stage)

Under requirements set by:

Auburn University

Collaborative Institutional Training Initiative
101 NE 3rd Avenue, Suite 320
Fort Lauderdale, FL 33301 US
www.citiprogram.org

Verify at www.citiprogram.org/verify/?wde865f8f-776d-48bc-8891-0af081417bca-25863903

CITI PROGRAM

Completion Date 08-Nov-2021
Expiration Date 07-Nov-2024
Record ID 25863898

This is to certify that:

Chang Ren

Not valid for renewal of certification through CME.

Has completed the following CITI Program course:

IRB Additional Modules
(Curriculum Group)
Research in Public Elementary and Secondary Schools - SBE
(Course Learner Group)
1 - Basic Course
(Stage)

Under requirements set by:

Auburn University

CITI
Collaborative Institutional Training Initiative
101 NE 3rd Avenue, Suite 320
Fort Lauderdale, FL 33301 US
www.citiprogram.org

Verify at www.citiprogram.org/verify/?w5e9636d9-6f71-48d1-8334-da8d8e91d5ea-25863898

CITI PROGRAM

Completion Date	15-May-2023
Expiration Date	15-May-2026
Record ID	55875700

This is to certify that:

Chang Ren

Has completed the following CITI Program course:

Not valid for renewal of certification through CME.

IRB Additional Modules
(Curriculum Group)
Internet Research - SBE
(Course Learner Group)
1 - Basic Course
(Stage)

Under requirements set by:

Auburn University

CITI
Collaborative Institutional Training Initiative

101 NE 3rd Avenue, Suite 320
Fort Lauderdale, FL 33301 US
www.citiprogram.org

Verify at www.citiprogram.org/verify/?wffeb5bfc-0bbf-4ada-bbcf-4590d25ee412-55875700

CITI PROGRAM

Completion Date 24-Jan-2023
Expiration Date 24-Jan-2026
Record ID 50321897

This is to certify that:

Chang Ren

Has completed the following CITI Program course:

Not valid for renewal of certification through CME.

Responsible Conduct of Research
(Curriculum Group)
AU Basic RCR Training for ALL Faculty, Staff, Postdocs, and Students
(Course Learner Group)
1 - RCR
(Stage)

Under requirements set by:

Auburn University

CITI
Collaborative Institutional Training Initiative
101 NE 3rd Avenue, Suite 320
Fort Lauderdale, FL 33301 US
www.citiprogram.org

Verify at www.citiprogram.org/verify/?w9412a3d2-2018-4f26-9609-01c0babda158-50321897

CITI PROGRAM

Completion Date 18-Jan-2022
Expiration Date 17-Jan-2025
Record ID 42389702

This is to certify that:

Cheryl Seals

Has completed the following CITI Program course:

Not valid for renewal of certification through CME.

IRB # 2 Social and Behavioral Emphasis - AU Personnel - Basic/Refresher
(Curriculum Group)

IRB # 2 Social and Behavioral Emphasis - AU Personnel
(Course Learner Group)

1 - Basic Course
(Stage)

Under requirements set by:

Auburn University

CITI
Collaborative Institutional Training Initiative

www.citiprogram.org/verify/?w17198472-33b9-4149-9b8d-37254afd1268-42389702

CITI PROGRAM

Completion Date	23-Sep-2021
Expiration Date	22-Sep-2024
Record ID	42389704

This is to certify that:

Cheryl Seals

Has completed the following CITI Program course:

Not valid for renewal of certification through CME.

IRB Additional Modules
(Curriculum Group)
History and Ethical Principles - SBE
(Course Learner Group)
1 - Basic Course
(Stage)

Under requirements set by:

Auburn University

CITI
Collaborative Institutional Training Initiative

www.citiprogram.org/verify/?wcabe5863-d5b4-4f39-a42b-5b18db20c4d4-42389704

Completion Date 27-Sep-2021
Expiration Date 26-Sep-2024
Record ID 42389703

This is to certify that:

Cheryl Seals

Has completed the following CITI Program course:

Not valid for renewal of certification through CME.

IRB Additional Modules
(Curriculum Group)

Conflicts of Interest in Research Involving Human Subjects
(Course Learner Group)

1 - Basic Course
(Stage)

Under requirements set by:

Auburn University

Collaborative Institutional Training Initiative

www.citiprogram.org/verify/?w80c2506c-69bc-496a-9e12-9fdaf9438f1e-42389703

CITI PROGRAM

Completion Date 16-Jan-2022
Expiration Date 15-Jan-2027
Record ID 28084134

This is to certify that:

Cheryl Seals

Has completed the following CITI Program course:

Not valid for renewal of certification through CME.

Responsible Conduct of Research for Social and Behavioral
(Curriculum Group)
Social, Behavioral and Education Sciences RCR
(Course Learner Group)
2 - RCR Refresher
(Stage)

Under requirements set by:

Auburn University

CITI
Collaborative Institutional Training Initiative

www.citiprogram.org/verify/?w4491791b-7b1c-42c2-8904-14b05eb9bce3-28084134

Revised 10/18/2022

AUBURN UNIVERSITY HUMAN RESEARCH PROTECTION PROGRAM (HRPP)

EXEMPT REVIEW APPLICATION

For assistance, contact: **The Office of Research Compliance (ORC)**
Phone: **334-844-5966** E-Mail: **IRBAdmin@auburn.edu** Web Address: **http://www.auburn.edu/research/vpr/ohs**
Submit completed form and supporting materials as one PDF through the IRB Submission Page
Hand written forms are not accepted. Where links are found hold down the control button (Ctrl) then click the link..

1. Project Identification **Today's Date: June 7, 2023**

Anticipated start date of the project: June 15, 2023 **Anticipated duration of project: 1 Year**

a. **Project Title: Enhancing User Experience through improving the User Interface of phonetics tools and studies on phone-level ASR-based automation through deep learning techniques**

b. **Principal Investigator (PI): Chang Ren** Degree(s): Click or tap here to enter text.
Rank/Title: Graduate Student Department/School: Computer Science and Software Engineering
Role/responsibilities in this project: **PI**
Preferred Phone Number: **3342755077** AU Email: czr0049@auburn.edu

Faculty Advisor Principal Investigator (if applicable): Cheryl Seals
Rank/Title: Professor Department/School: Computer Science and Software Engineering
Role/responsibilities in this project: **Co-I**
Preferred Phone Number: **3348446319** AU Email: sealscd@auburn.edu

Department Head: Hari Narayanan Department/School: Computer Science and Software Engineering
Preferred Phone Number: **3348446312** AU Email: naraynh@auburn.edu
Role/responsibilities in this project: Click or tap here to enter text.

c. **Project Key Personnel –** Identify all key personnel who will be involved with the conduct of the research and describe their role in the project. Role may include design, recruitment, consent process, data collection, data analysis, and reporting. (To determine key personnel, see decision tree**).** *Exempt determinations are made by individual institutions; reliance on other institutions for exempt determination is not feasible. Non-AU personnel conducting exempt research activities must obtain approval from the IRB at their home institution.*

Key personnel are required to maintain human subjects training through CITI. Only for EXEMPT level research is documentation of completed CITI training NO LONGER REQUIRED to be included in the submission packet. NOTE however, **the IRB will perform random audits of CITI training records to confirm** reported training courses and expiration dates. Course title and expiration dates are shown on training certificates.

Name: Chang Ren Degree(s): Click or tap here to enter text.
Rank/Title: Graduate Student Department/School: Computer Science and Software Engineering
Role/responsibilities in this project: PI Ren will be responsible for conducting research, designing and implementing the necessary applications, developing protocols and surveys, carrying out experiments, recruiting and interacting with participants.
- AU affiliated? ☒ Yes ☐ No If no, name of home institution: Click or tap here to enter text.
- Plan for IRB approval for non-AU affiliated personnel? Click or tap here to enter text.
- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No
- If yes, briefly describe the potential or real conflict of interest: Click or tap here to enter text.
- Completed required CITI training? ☒ Yes ☐ No If NO, complete the appropriate CITI basic course and update the revised Exempt Application form.

The Auburn University Institutional Review Board has approved this Document for use from 06/06/2023 to ------------
Protocol # 23-284 EX 2306

- If YES, choose course(s) the researcher has completed: Conflicts of Interest in Research Involving Human Subjects(2024), Defining Research with Human Subjects(2024), History and Ethical Principles(2024), IRB # 2 Social and Behavioral Emphasis(2024), Research in Public Elementary and Secondary Schools (2025), Internet Research(2026), AU Basic RCR Training for ALL Faculty, Staff, Postdocs, and Students (2026).

Name: Dr. Cheryl Seals Degree(s): Click or tap here to enter text.
Rank/Title: Professor Department/School: Computer Science and Software Engineering
Role/responsibilities in this project: Co-I Seals will assist with research prepare protocols, surveys and assist experimental trials and meet with participants.
- AU affiliated? ☒ Yes ☐ No If no, name of home institution: Click or tap here to enter text.
- Plan for IRB approval for non-AU affiliated personnel? Click or tap here to enter text.
- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☒ No
- If yes, briefly describe the potential or real conflict of interest: Click or tap here to enter text.
- Completed required CITI training? ☐ Yes ☐ No If NO, complete the appropriate CITI basic course and update the revised EXEMPT application form.
- If YES, choose course(s) the researcher has completed: Conflicts of Interest in Research Involving Human Subjects (2024), IRB # 2 Social and Behavioral Emphasis (2025), History and Ethical Principles (2024), Responsible Conduct of Research for Social and Behavioral (2027).

Name: Click or tap here to enter text. Degree(s): Click or tap here to enter text.
Rank/Title: Choose Rank/Title Department/School: Choose Department/School
Role/responsibilities in this project: Click or tap here to enter text.
- AU affiliated? ☐ Yes ☐ No If no, name of home institution: Click or tap here to enter text.
- Plan for IRB approval for non-AU affiliated personnel? Click or tap here to enter text.
- Do you have any known competing financial interests, personal relationships, or other interests that could have influence or appear to have influence on the work conducted in this project? ☐ Yes ☐ No
- If yes, briefly describe the potential or real conflict of interest: Click or tap here to enter text.
- Completed required CITI training? ☐ Yes ☐ No If NO, complete the appropriate CITI basic course and update the revised EXEMPT application form.
- If YES, choose course(s) the researcher has completed: Choose a course Expiration Date
Choose a course Expiration Date

d. Funding Source – Is this project funded by the investigator(s)? Yes ☒ No ☐
Is this project funded by AU? Yes ☐ No ☒ If YES, identify source Click or tap here to enter text.
Is this project funded by an external sponsor? Yes ☐ No ☒ If YES, provide name of sponsor, type of sponsor (governmental, non-profit, corporate, other), and an identification number for the award.
Name: Click or tap here to enter text. Type: Click or tap here to enter text. Grant #: Click or tap here to enter text.

e. List other AU IRB-approved research projects and/or IRB approvals from other institutions that are associated with this project. Describe the association between this project and the listed project(s):
Click or tap here to enter text.

2. Project Summary

a. Does the study *TARGET* any special populations? Answer YES or NO to all.

Minors (under 18 years of age; if minor participants, at least 2 adults must be present during all research procedures that include the minors)	Yes ☐ No ☒
Auburn University Students	Yes ☒ No ☐
Pregnant women, fetuses, or any products of conception	Yes ☐ No ☒

Prisoners or wards (unless incidental, not allowed for Exempt research) Yes ☐ No ☒

Temporarily or permanently impaired Yes ☐ No ☒

b. Does the research pose more than minimal risk to participants? Yes ☐ No ☒

If YES, to question 2.b, then the research activity is NOT eligible for EXEMPT review. Minimal risk means that the probability and magnitude of harm or discomfort anticipated in the research is not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or test. 42 CFR 46.102(i)

c. Does the study involve any of the following? *If YES to any of the questions in item 2.c, then the research activity is NOT eligible for EXEMPT review.*

Procedures subject to FDA regulations (drugs, devices, etc.) Yes ☐ No ☒

Use of school records of identifiable students or information from instructors about specific students. Yes ☐ No ☒

Protected health or medical information when there is a direct or indirect link which could identify the participant. Yes ☐ No ☒

Collection of sensitive aspects of the participant's own behavior, such as illegal conduct, drug use, sexual behavior or alcohol use. Yes ☐ No ☒

d. Does the study include deception? Requires limited review by the IRB* Yes ☐ No ☒

3. MARK the category or categories below that describe the proposed research. Note the IRB Reviewer will make the final determination of the eligible category or categories.

☒ **1.** Research conducted in established or commonly accepted educational settings, involving normal educational practices. The research is not likely to adversely impact students' opportunity to learn or assessment of educators providing instruction. 104(d)(1)

☒ **2.** Research only includes interactions involving educational tests, surveys, interviews, public observation if at least ONE of the following criteria. (The research includes data collection only; may include visual or auditory recording; may NOT include intervention and only includes interactions). **Mark the applicable sub-category below (I, ii, or iii). 104(d)(2)**

☒ **(i)** Recorded information cannot readily identify the participant (directly or indirectly/ linked);
OR
- surveys and interviews: no children;
- educational tests or observation of public behavior: can only include children when investigators do not participate in activities being observed.

☒ **(ii)** Any disclosures of responses outside would not reasonably place participant at risk; **OR**

☐ **(iii)** Information is recorded with identifiers or code linked to identifiers and IRB conducts limited review; no children. **Requires limited review by the IRB.***

☐ **3.** Research involving Benign Behavioral Interventions (BBI)** through verbal, written responses including data entry or audiovisual recording from adult subjects who prospectively agree and ONE of the following criteria is met. (This research does not include children and does not include medical interventions. Research cannot have deception unless the participant prospectively agrees that they will be unaware of or misled regarding the nature and purpose of the research) **Mark the applicable sub-category below (A, B, or C).** 104(d)(3)(i)

☐ **(A)** Recorded information cannot readily identify the subject (directly or indirectly/ linked); **OR**

☐ **(B)** Any disclosure of responses outside of the research would not reasonably place subject at risk; **OR**

☐ **(C)** Information is recorded with identifies and cannot have deception unless participants prospectively agree. **Requires limited review by the IRB.***

☐ **4.** Secondary research for which consent is not required: use of identifiable information or identifiable bio-specimen that have been or will be collected for some other 'primary' or 'initial' activity, if one of the following criteria is met. Allows retrospective and prospective secondary use. **Mark the applicable sub-category below (i, ii, iii, or iv).** 104 (d)(4)

☐ **(i)** Bio-specimens or information are publicly available;

☐ **(ii)** Information recorded so subject cannot readily be identified, directly or indirectly/linked investigator does not contact subjects and will not re-identify the subjects; **OR**

☐ **(iii)** Collection and analysis involving investigators use of identifiable health information when us is regulated by HIPAA "health care operations" or "research" or "public health activities and purposes" (does not include bio-specimens (only PHI and requires federal guidance on how to apply); **OR**

☐ **(iv)** Research information collected by or on behalf of federal government using government generated or collected information obtained for non-research activities.

☐ **5.** Research and demonstration projects which are supported by a federal agency/department AND designed to study and which are designed to study, evaluate, or otherwise examine: (i)public benefit or service programs; (ii) procedures for obtaining benefits or services under those programs; (iii) possible changes in or alternatives to those programs or procedures; or (iv) possible changes in methods or levels of payment for benefits or service under those programs. (must be posted on a federal web site). 104.5(d)(5) (must be posted on a federal web site)

☐ **6.** Taste and food quality evaluation and consumer acceptance studies, (i) if wholesome foods without additives and consumed or (ii) if a food is consumed that contains a food ingredient at or below the level and for a use found to be safe, or agricultural chemical or environmental contaminant at or below the level found to be safe, by the Food and Drug Administration or approved by the Environmental Protection Agency or the Food Safety and Inspection Service of the U.S. Department of Agriculture. The research does not involve prisoners as participants. 104(d)(6)

**Limited IRB review – the IRB Chair or designated IRB reviewer reviews the protocol to ensure adequate provisions are in place to protect privacy and confidentiality.*

***Category 3 – Benign Behavioral Interventions (BBI) must be brief in duration, painless/harmless, not physically invasive, not likely to have a significant adverse lasting impact on participants, and it is unlikely participants will find the interventions offensive or embarrassing.*

**** Exemption categories 7 and 8 require broad consent. The AU IRB has determined the regulatory requirements for legally effective broad consent are not feasible within the current institutional infrastructure. EXEMPT categories 7 and 8 will not be implemented at this time.*

4. Describe the proposed research including who does what, when, where, how, and for how long, etc.

a. Purpose

The purpose of this study is to enhance the user interface design and user experience of the Automated Phonetic Transcription Grading Tool (APTgt), an interactive web-based E-learning system designed to support communication disorders faculty in teaching phonetic transcription. The research aims to optimize the current designs by improving ease of use, aesthetics, and consistency in design, with the ultimate goal of providing a more efficient and user-friendly tool for linguistic instructors. The study specifically addresses the following questions: How can the user interface of APTgt be optimized to improve ease of use, aesthetics, and consistency in design? How can the efficiency of APTgt be increased to reduce the time required for instructors to use the software?

b. Participant population, including the number of participants and the rationale for determining number of participants to recruit and enroll. Note if the study enrolls minor participants, describe the process to ensure more than 1 adult is present during all research procedures which include the minor.

The participants will be linguistic professionals (i.e. advanced undergraduate, graduates or faculty in a field that is practiced in linguistic transcription) and usability experts who possess usability certification, have completed a course in usability, or have relevant experience as a usability researcher or member of technical staff). We will employ email and opportunistic sampling as well as snowball recruiting methods to determine the number of participants we would like to recruit.

c. Recruitment process. Address whether recruitment includes communications/interactions between study staff and potential participants either in person or online. *Submit a copy of all recruitment materials.*

We will recruit using email to linguistic instructors that we have worked with on prior projects, opportunistic sampling, and snowball recruiting. We will provide an example recruitment email in the IRB appendix and will send them a survey link: https://auburn.qualtrics.com/jfe/form/SV_2i3EltzcU7HCn5k

d. Consent process including how information is presented to participants, etc.

The information letter will be used for consent.

e. Research procedures and methodology

To evaluate the effectiveness of the enhancements, participants will be provided with two versions of the APTgt application—an existing benchmark version and an experimental version with the proposed improvements. The study will involve conducting a comparative usability evaluation. Participants will first complete a pre-questionnaire to gather their initial perceptions and expectations of the application (5-10 minutes). They will then use both versions of the application for a duration of 20-40 minutes. Finally, the session will conclude with a post-questionnaire to gather feedback on aesthetics, ease of use, satisfaction, and overall usability of the application (5-10 minutes).

f. Anticipated time per study exercise/activity and total time if participants complete all study activities.

Duration of time needed to complete the pre-survey is 5 - 10 minutes and post-survey is 5 - 10 minutes. The anticipated time it will take to use the applications is 20-40 minutes. The anticipated total time to complete all study activities is 30 – 60 minutes.

g. Location of the research activities.

The research will be done online.

h. Costs to and compensation for participants? If participants will be compensated describe the amount, type, and process to distribute.

There will be no compensations.

i. Non-AU locations, site, institutions. *Submit a copy of agreements/IRB approvals.*
Click or tap here to enter text.

j. Describe how results of this study will be used (presentation? publication? thesis? dissertation?)
The result of this study will be used for a Ph.D. dissertation.

k. Additional relevant information.
Click or tap here to enter text.

5. Waivers
Check applicable waivers and describe how the project meets the criteria for the waiver.

☐ Waiver of Consent (Including existing de-identified data)

☒ Waiver of Documentation of Consent (Use of Information Letter, rather than consent form requiring signatures)

☐ Waiver of Parental Permission (in Alabama, 18 years-olds may be considered adults for research purposes)
https://sites.auburn.edu/admin/orc/irb/IRB 1 Exempt and Expedited/11-113 MR 1104 Hinton Renewal 2021-1.pdf

a. Provide the rationale for the waiver request.
We have updated the Informed Consent to an Information Letter by eliminating unnecessary signatures. The data collected will be anonymous for those who complete pre-survey only. For those who choose to continue with the experience, we will assign a code that links their pre-to-post survey results. (i.e. the data will be collected and stored on Qualtrics servers and Auburn Box to ensure confidentiality. Access to the survey results will be limited to the Primary Investigator and Co-Investigator only.).

6. Describe the process to select participants/data/specimens. If applicable, include gender, race, and ethnicity of the participant population.

The participants will be linguistic professionals and usability experts who are undergraduate, graduate students or faculty at Auburn University and are above 18 years old. We will recruit participants of all genders and distribute the surveys online via Qualtrics.

7. Risks and Benefits

7a. Risks - Describe why none of the research procedures would cause a participant either physical or psychological discomfort or be perceived as discomfort above and beyond what the person would experience in daily life (minimal risk).

There are no risks associated with this research.

7b. Benefits – Describe whether participants will benefit directly from participating in the study. If yes, describe the benefit. And, describe generalizable benefits resulting from the study.

The participation in the study will not provide direct personal benefits to the participants. However, participating students may gain advanced knowledge in the area of User Interface Design, while faculty members may benefit from enhanced teaching experiences using the E-learning system.

8. Describe the provisions to maintain confidentiality of data, including collection, transmission, and storage. Identify platforms used to collect and store study data. *For EXEMPT research, the AU IRB recommends AU BOX or using an AU issued and encrypted device. If a data collection form will be used, submit a copy.*

The survey data will be securely stored on Qualtrics servers and will be stored on Auburn Box. Only Primary Investigator and Co-Investigators have access to the survey results.

a. If applicable, submit a copy of the data management plan or data use agreement.

9. Describe the provisions included in the research to protect the privacy interests of participants (e.g., others will not overhear conversations with potential participants, individuals will not be publicly identified or embarrassed).

Participant data will be collected anonymously, with no connection to identifying information. The survey data will be securely stored on Qualtrics servers and Auburn Box, and only the Primary Investigator and Co-Investigators will have access to the survey results.

10. Does this research include purchase(s) that involve technology hardware, software or online services?

☐ **YES** ☒ **NO**

If YES:

A. Provide the name of the product Click or tap here to enter text.
and the manufacturer of the product Click or tap here to enter text.

B. Briefly describe use of the product in the proposed human subject's research.
Click or tap here to enter text.

C. To ensure compliance with AU's Electronic and Information Technology Accessibility Policy, contact AU IT Vendor Vetting team at vetting@auburn.edu to learn the vendor registration process (prior to completing the purchase).

D. Include a copy of the documentation of the approval from AU Vetting with the revised submission.

11. Additional Information and/or attachments.

In the space below, provide any additional information you believe may help the IRB review of the proposed research. If attachments are included, list the attachments below. Attachments may include recruitment materials, consent documents, site permissions, IRB approvals from other institutions, data use agreements, data collection form, CITI training documentation, etc.

Information Letter
Recruitment Email
Printed version of online survey. (https://auburn.qualtrics.com/jfe/form/SV_2i3EltzcU7HCn5k)
Chang Ren Citi Training
Cheryl Seals Citi Training

Required Signatures *(If a student PI is identified in item 1.a, the EXEMPT application must be re-signed and updated at every revision by the student PI and faculty advisor. The signature of the department head is required only on the initial submission of the EXEMPT application, regardless of PI. Staff and faculty PI submissions require the PI signature on all* *version, the department head signature on the original submission)*

Signature of Principal Investigator: Chang Ren **Date:** 06/07/2023

Signature of Faculty Advisor (If applicable): [signature] **Date:** 06/08/2023

Signature of Dept. Head: N. Hung Nguyen **Date:** 5/26/23

Version Date: **6/7/2023**

AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE AND SOFTWARE ENGINEERING

(NOTE: DO NOT AGREE TO PARTICIPATE UNLESS IRB APPROVAL INFORMATION WITH CURRENT DATES HAS BEEN ADDED TO THIS DOCUMENT.)

INFORMATION LETTER
for a Research Study entitled
"Enhancing User Experience through improving the User Interface of phonetics tools and studies on phone-level ASR-based automation through deep learning techniques"

You are invited to participate in a research study to enhance the user interface and user experience of the Automated Phonetic Transcription Grading Tool (APTgt), an interactive web-based E-learning system designed to support communication disorders faculty in teaching phonetic transcription. The goal is to optimize the current designs by improving ease of use, aesthetics, and design consistency. The study is being conducted by Chang Ren, PhD candidate, under the direction of Cheryl Seals, Charles W. Barkley Professor in the Auburn University Department of Computer Science and Software Engineering. You are invited to participate because you are Linguistic professionals or Usability experts and are age 19 or older.

What will be involved if you participate?
Your participation is completely voluntary. If you decide to participate in this research study, you will be asked to complete an online survey about your educational background, qualifications, specific areas of expertise, and previous experience with E-learning systems or user interface design. Your total time commitment will be approximately 5 – 10 minutes.
To continue participating in the follow-up experiment and task activities, you can express your interest by providing your contact email on a separate form that does not collect any identifying information. After completing the initial survey, you will be directed to an online form where you can enter your email address. If you do not wish to be contacted, you can simply close the tab containing the form. Participants who indicate their willingness to be selected will be contacted for task activities and an online post-questionnaire once we are ready to gather feedback on application evaluation. During this process, participants will engage in usability evaluation of the application and complete a post-questionnaire. The estimated time required to complete task activities using both versions of the application is anticipated to be between 20 to 40 minutes. Post- Questionnaire take approximately 5–10 minutes.

Are there any risks or discomforts? There are no risks or discomforts involved.

Are there any benefits to yourself or others?
There are no personal benefits directly associated with participating in this research. However, if you participate in this study, you can expect to contribute your expertise to enhance the user interface and user experience of phonetic E-learning system. The involvement will directly influence the system's improvement, making it more efficient and user-friendly for linguistic instructors. Additionally, participants will gain hands-on experience in usability evaluation, enhancing your professional skills and expertise in user-centered design. Your contributions will shape the future of E-learning in communication disorders, benefiting educators and students in the field.

The Auburn University Institutional Review Board has approved this Document for use from
06/06/2023 to -------------
Protocol # 23-284 EX 2306

Will you receive compensation for participating? There is no compensation for participation.
Are there any costs? There are no costs to participate in the research.

If you change your mind about participating, you can withdraw at any time by stopping the survey. Then the data will not be collected or recorded.

Any data obtained in connection with this study will remain anonymous and confidential. The study will take place via an online survey. Collected data with Auburn Qualtrics will be stored on Auburn Box. Only the investigators of this research will have access to the data obtained.

If you have questions about this study, please contact Chang Ren at czr0049@auburn.edu or Dr. Cheryl Seals at sealscd@auburn.edu.

If you have questions about your rights as a research participant, you may contact the Auburn University Office of Research Compliance or the Institutional Review Board by phone (334) 844-5966 or e-mail at IRBadmin@auburn.edu or IRBChair@auburn.edu.

HAVING READ THE INFORMATION ABOVE, YOU MUST DECIDE IF YOU WANT TO PARTICIPATE IN THIS RESEARCH PROJECT. IF YOU DECIDE TO PARTICIPATE, PLEASE CLICK ON THE LINK BELOW.
YOU MAY PRINT A COPY OF THIS LETTER TO KEEP.

Chang Ren ____________ 06/07/2023
Investigator Date

Chang Ren
Printed Name

Cheryl P. Seals ____________ 06/08/2023
Co-Investigator Date

Dr. Cheryl Seals
Printed Name

LINK TO SURVEY

Version Date (date document created): 06/07/2023

The Auburn University Institutional Review Board has approved this Document for use from 06/06/2023 to -------------
Protocol # 23-284 EX 2306

RECRUITMENT EMAIL

Dear potential participants,

Thank you for taking the time to read this email. My name is Chang Ren, a Ph.D. candidate studying Computer Science and Software Engineering at Auburn University. Under the guidance of my advisor, Dr. Cheryl Seals, Professor in the Auburn University Department of Computer Science and Software Engineering, we are conducting a survey targeting linguistic professionals and usability experts to evaluate the optimized user interface design and user experience of the Automated Phonetic Transcription Grading Tool (APTgt), an online E-learning system designed to support communication disorders faculty.

Your participation is completely voluntary. If you decide to participate in this research study, you will be asked to complete an online survey about your educational background, qualifications, specific areas of expertise, and previous experience with E-learning systems or user interface design. Then we are specifically interested in gathering opinion and feedback on ease of use, aesthetics, consistency in design, and overall usability of the application. Your insights will be instrumental in optimizing the user interface design to provide a more efficient and user-centered experience for linguistic instructors. Your commitment time for this survey will be approximately 5 - 10 minutes.

If you have questions about this study, please contact Chang Ren at czr0049@auburn.edu or Dr. Cheryl Seals at sealscd@auburn.edu. You can find the full Information Letter attached below.

SURVEY LINK

https://auburn.qualtrics.com/jfe/form/SV_2i3EltzcU7HCn5k

Thank you again for considering participation in our research. Your input and expertise are greatly appreciated.

Best regards,
Chang Ren
Ph.D. Candidate
Department of Computer Science and Software Engineering
Auburn University, Auburn, AL

The Auburn University Institutional Review Board has approved this Document for use from
06/06/2023 to -------------
Protocol # 23-284 EX 2306

PRE-SURVEY

Survey Completion

0% 100%

AUBURN UNIVERSITY

Hello! **You are invited to participate in a research study** to enhance the user interface and user experience of the Automated Phonetic Transcription Grading Tool (APTgt), an interactive web-based E-learning system designed to support communication disorders faculty in teaching phonetic transcription. The goal is to optimize the current designs by improving ease of use, aesthetics, and design consistency. The study is being conducted by Chang Ren, PhD candidate, under the direction of Cheryl Seals, Charles W. Barkley Professor in the Auburn University Department of Computer Science and Software Engineering. You are invited to participate because you are Linguistic professionals or Usability experts and are age 19 or older.

What will be involved if you participate? Your participation is completely voluntary. If you decide to participate in this research study, you will be asked to complete an online survey about your educational background, qualifications, specific areas of expertise, and previous experience with E-learning systems or user interface design. Your total time commitment will be approximately 5 – 10 minutes. To continue participating in the follow-up experiment and task activities, you can express your interest by providing your contact email on a separate form that does not collect any identifying information. After completing the initial survey, you will be directed to an online form where you can enter your email address. If you do not wish to be contacted, you can close the tab containing the form. Participants who indicate their willingness to be selected will be contacted for task activities and an online post-questionnaire once we are ready to gather feedback on the application evaluation. During this process, participants will engage in a usability evaluation of the application and complete a post-questionnaire. Post-Questionnaire takes approximately 5–10 minutes.

Are there any risks or discomforts? There are no risks or discomforts involved.

Version Date (date document created): 05/25/2023

The Auburn University Institutional Review Board has approved this Document for use from 06/06/2023 to -------------

Protocol # 23-284 EX 2306

Are there any benefits to yourself or others? If you participate in this study, you can expect to contribute your expertise to enhance the user interface and user experience of the phonetic E-learning system. The involvement will directly influence the system's improvement, making it more efficient and user-friendly for linguistic instructors. Additionally, participants will gain hands-on experience in usability evaluation, enhancing their professional skills and expertise in user-centered design. Your contributions will shape the future of E-learning in communication disorders, benefiting educators and students in the field.

Will you receive compensation for participating? There is no compensation for participation.

Are there any costs? There are no costs to participate in the research.

If you change your mind about participating, you can withdraw at any time by stopping the survey. Then the data will not be collected or recorded.

Any data obtained in connection with this study will remain anonymous and confidential. The study will take place via an online survey. Collected data with Auburn Qualtrics will be stored on Auburn Box. Only the investigators of this research will have access to the data obtained.

If you have questions about this study, please contact Chang Ren at czr0049@auburn.edu or Dr. Cheryl Seals at sealscd@auburn.edu.

If you have questions about your rights as a research participant, you may contact the Auburn University Office of Research Compliance or the Institutional Review Board by phone (334) 844-5966 or e-mail at IRBadmin@auburn.edu or IRBChair@auburn.edu.

HAVING READ THE INFORMATION ABOVE, YOU MUST DECIDE IF YOU WANT TO PARTICIPATE IN THIS RESEARCH PROJECT. IF YOU DECIDE TO PARTICIPATE, PLEASE CLICK ON THE LINK BELOW. YOU MAY PRINT A COPY OF THIS LETTER TO KEEP.

››

Powered by Qualtrics

The Auburn University Institutional Review Board has approved this Document for use from
06/06/2023 to -------------
Protocol # 23-284 EX 2306

AUBURN UNIVERSITY

What is your gender?

- ○ Female
- ○ Male
- ○ Not Listed

How old are you?

- ○ 19 - 24
- ○ 25 - 34
- ○ 35 +

What is your current role?

- ○ Undergradute Student
- ○ Graduate Student
- ○ Professional

What are your areas of expertise? *Select all that apply.*

- ☐ Linguistics
- ☐ Usability
- ☐ Other (specify)

How many years of experience do you have in your field?

- ○ 0 - 2 years
- ○ 3 - 5 years
- ○ 6 - 10 years
- ○ 10 + years

The Auburn University Institutional Review Board has approved this Document for use from 06/06/2023 to ------------
Protocol # 23-284 EX 2306

Do you have prior experience working in an E-learning system?

- ◯ Yes
- ◯ No

How familiar you are with E-learning system?

- ◯ Very familiar, I have extensive experience using E-learning systems.
- ◯ Moderately familiar, I have some experience using E-learning systems.
- ◯ Somewhat familiar, I have limited experience using E-learning systems.

what is your most frequently used E-learning system? *Select all that apply.*

- ☐ Canvas
- ☐ Moodle
- ☐ Blackboard Learn
- ☐ Coursera
- ☐ Udemy
- ☐ Other (specify)

Are there any specific challenges you have encountered when using E-learning systems? *Select all that apply.*

- ☐ Unappealing
- ☐ Inconsistent design
- ☐ Cluttered/complex interface
- ☐ Poor use of colors/typography/layout.
- ☐ Insufficient user control
- ☐ Navigation difficulties
- ☐ Lack of feedback
- ☐ Other (specify)

Do you have prior experience working in an E-learning system?

- ◯ Yes
- ◯ No

What are your goals for using APTgt? *Select all that apply.*

- ☐ Manage online courses/exams in linguistics.
- ☐ Improve my skills in phonetic transcription.
- ☐ Conduct research in communication disorders.
- ☐ Other (specify)

How often do you access APTgt?

- ◯ Daily
- ◯ A few times per week
- ◯ A few times per month
- ◯ Once or twice a semester or less often

Are there any specific challenges you have encountered when using APTgt? *Select all that apply.*

- ☐ Unappealing
- ☐ Inconsistent design
- ☐ Cluttered/complex interface
- ☐ Poor use of colors/typography/layout.
- ☐ Insufficient user control
- ☐ Navigation difficulties
- ☐ Lack of feedback
- ☐ Other (specify)

How likely are you to use APTgt again in the future?

- ○ Extremely unlikely
- ○ Somewhat unlikely
- ○ Neither likely nor unlikely
- ○ Somewhat likely
- ○ Extremely likely

‹‹ ››

Have you been involved in usability evaluations in the past?

- ◉ Yes
- ○ No

How many usability evaluations have you been involved in?

- ○ 1-5
- ○ 6-10
- ○ 11-15
- ○ 16-20
- ○ 21+

Which usability evaluation methods are you most experienced in? *Select all that apply.*

- ☐ User interviews
- ☐ User testing
- ☐ Heuristic evaluations
- ☐ Cognitive walkthroughs
- ☐ Surveys/questionnaires
- ☐ Other (specify)

In which domains have you primarily experienced in usability evaluations? *Select all that apply.*

- [ ] Software
- [ ] Web application
- [ ] Mobile
- [ ] Education
- [ ] Healthcare
- [ ] E-commerce
- [ ] Gaming
- [ ] Manufacturing
- [ ] Retail
- [ ] Other (please specify)

Have you involved in usability evaluations for E-learning systems in the past?

- ○ Yes, multiple times
- ○ Yes, once or twice
- ○ No.

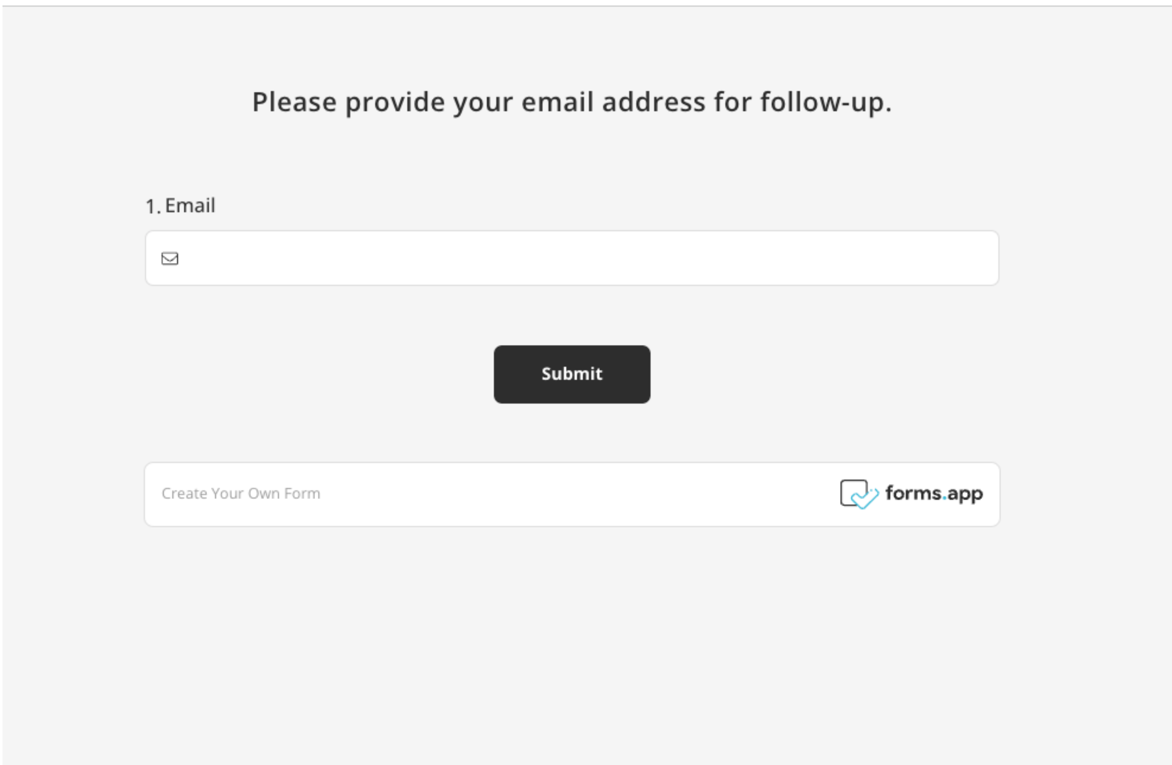
Thank you so much for your time spent taking this survey.

If you would like to be contacted in the future for feedback or evaluation of the system, please proceed to the end of the survey where you will be redirected to an online form to enter your email address.

If you do not wish to be contacted, please continue to the end of the survey. After you are redirected, please simply close the webpage to opt out.

<< >>

CONTACT FORM



Please provide your email address for follow-up.

1. Email

Submit

Create Your Own Form

forms.app

The Auburn University Institutional Review Board has approved this Document for use from
06/06/2023 to -------------
Protocol # 23-284 EX 2306

POST SURVEY

AUBURN
UNIVERSITY

Please select your role or select "Other."

- ○ Linguistic Professional
- ○ Usability Expert
- ○ Other (please specify)

Overall, what's your experience been with the APTgt v2.0?

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
The application is more visually pleasing than old version.	○	○	○	○	○
The application is easier to use than old version.	○	○	○	○	○
The navigation is more efficient than old version.	○	○	○	○	○
I am more satisfied with the application than the old version.	○	○	○	○	○

In the scale of 1-5, please rate the APTgt v2.0 in terms of the following aspects?

	Excellent	Very Good	Good	Poor	Very poor
Layout	○	○	○	○	○
Use of color	○	○	○	○	○
Consistency	○	○	○	○	○
Easy to use	○	○	○	○	○
Navigation	○	○	○	○	○

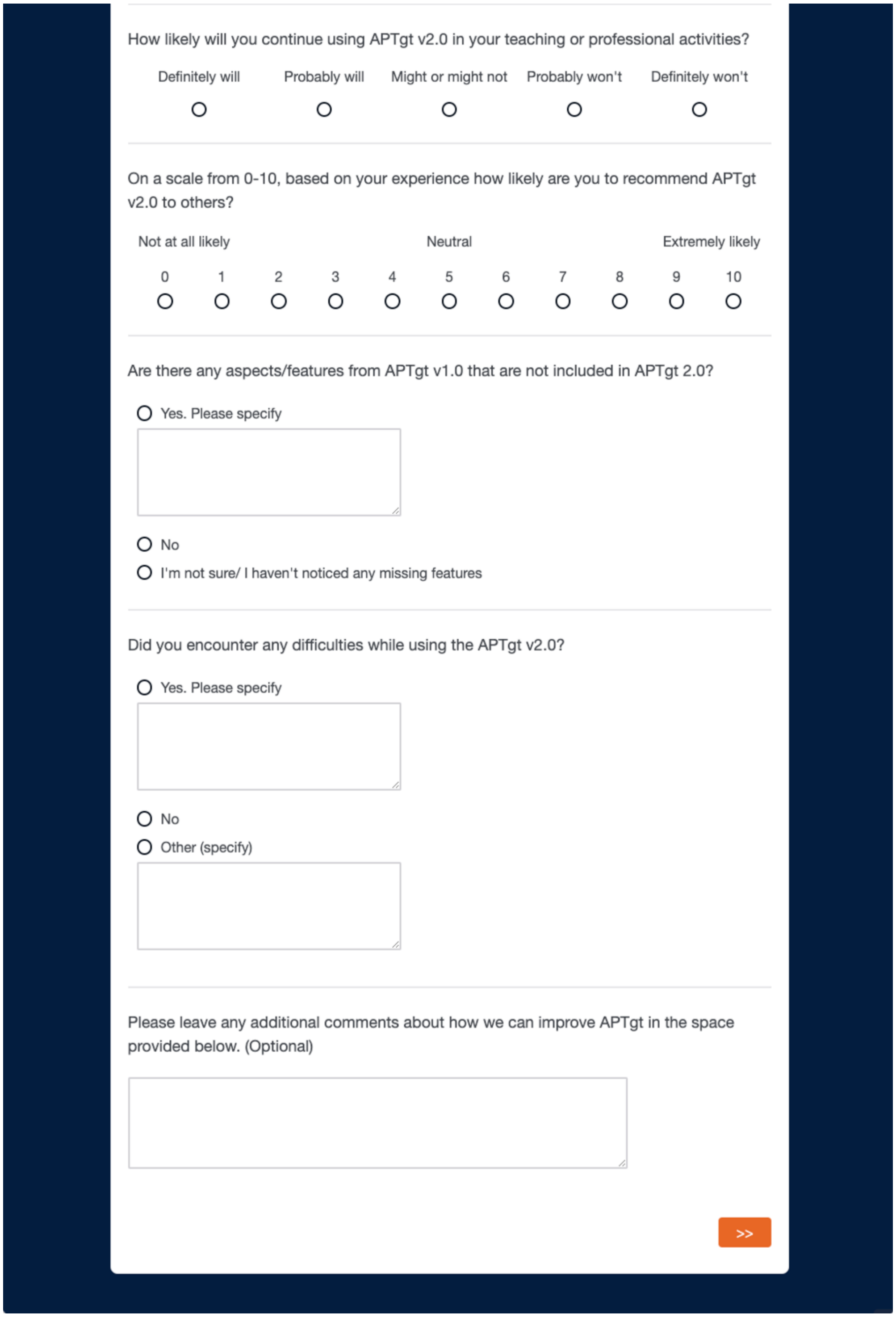
The APTgt v2.0 help to complete tasks more efficiently than the old version?

Strongly Agree	Somewhat agree	Neutral	Somewhat disagree	Strongly disagree
○	○	○	○	○

How would you rate the overall usability of APTgt v2.0 compared to the old version?

Much more usable	Somewhat more usable	Same	Somewhat less usable	Much less usable
○	○	○	○	○

The Auburn University Institutional Review Board has approved this Document for use from
06/06/2023 to -------------
Protocol # 23-284 EX 2306



How likely will you continue using APTgt v2.0 in your teaching or professional activities?

Definitely will	Probably will	Might or might not	Probably won't	Definitely won't
○	○	○	○	○

On a scale from 0-10, based on your experience how likely are you to recommend APTgt v2.0 to others?

Not at all likely — Neutral — Extremely likely

0	1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○	○

Are there any aspects/features from APTgt v1.0 that are not included in APTgt 2.0?

- ○ Yes. Please specify
- ○ No
- ○ I'm not sure/ I haven't noticed any missing features

Did you encounter any difficulties while using the APTgt v2.0?

- ○ Yes. Please specify
- ○ No
- ○ Other (specify)

Please leave any additional comments about how we can improve APTgt in the space provided below. (Optional)

>>

The following survey questions will be added to the post survey for more comprehensive insights from evaluators.

1. Overall, I am satisfied with the ease of completing the tasks in this scenario. (Strongly disagree – Strongly agree)
2. Overall, I am satisfied with the amount of time it took to complete the tasks in this scenario. (Strongly disagree – Strongly agree)
3. It's user-friendly. (Strongly disagree – Strongly agree)
4. The interface of this system is pleasant
5. I can use it without written instructions. (Strongly disagree – Strongly agree)
6. It is pleasant to use. (Strongly disagree – Strongly agree)
7. It works the way I want it to work. (Strongly disagree – Strongly agree)
8. My interaction with the system would be clear and understandable (unlikely – likely)

The Auburn University Institutional Review Board has approved this Document for use from
06/06/2023 to -------------
Protocol # 23-284 EX 2306

Completion Date 24-Feb-2021
Expiration Date 24-Feb-2024
Record ID 25863904

This is to certify that:

Chang Ren

Has completed the following CITI Program course:

Not valid for renewal of certification through CME.

IRB Additional Modules
(Curriculum Group)
Conflicts of Interest in Research Involving Human Subjects
(Course Learner Group)
1 - Basic Course
(Stage)

Under requirements set by:

Auburn University

Collaborative Institutional Training Initiative
101 NE 3rd Avenue, Suite 320
Fort Lauderdale, FL 33301 US
www.citiprogram.org

Verify at www.citiprogram.org/verify/?w4fe0b188-9ac4-4763-9f29-24a9ce366e50-25863904

CITI PROGRAM

Completion Date 24-Feb-2021
Expiration Date 24-Feb-2024
Record ID 25863906

This is to certify that:

Chang Ren

Has completed the following CITI Program course:

Not valid for renewal of certification through CME.

IRB Additional Modules
(Curriculum Group)
Defining Research with Human Subjects - SBE
(Course Learner Group)
1 - Basic Course
(Stage)

Under requirements set by:

Auburn University

CITI
Collaborative Institutional Training Initiative
101 NE 3rd Avenue, Suite 320
Fort Lauderdale, FL 33301 US
www.citiprogram.org

Verify at www.citiprogram.org/verify/?w41755bce-1025-4028-800a-a51df2c4d5eb-25863906

CITI PROGRAM

Completion Date 24-Feb-2021
Expiration Date 24-Feb-2024
Record ID 39055407

This is to certify that:

Chang Ren

Not valid for renewal of certification through CME.

Has completed the following CITI Program course:

IRB Additional Modules
(Curriculum Group)
History and Ethical Principles - SBE
(Course Learner Group)
1 - Basic Course
(Stage)

Under requirements set by:

Auburn University

CITI
Collaborative Institutional Training Initiative

101 NE 3rd Avenue, Suite 320
Fort Lauderdale, FL 33301 US
www.citiprogram.org

Verify at www.citiprogram.org/verify/?w8238f834-ed39-4849-8a98-31c1bf2c2eca-39055407

CITI PROGRAM

Completion Date 24-Feb-2021
Expiration Date 24-Feb-2024
Record ID 25863903

This is to certify that:

Chang Ren

Not valid for renewal of certification through CME.

Has completed the following CITI Program course:

IRB # 2 Social and Behavioral Emphasis - AU Personnel - Basic/Refresher
(Curriculum Group)
IRB # 2 Social and Behavioral Emphasis - AU Personnel
(Course Learner Group)
1 - Basic Course
(Stage)

Under requirements set by:

Auburn University

CITI
Collaborative Institutional Training Initiative

101 NE 3rd Avenue, Suite 320
Fort Lauderdale, FL 33301 US
www.citiprogram.org

Verify at www.citiprogram.org/verify/?wde865f8f-776d-48bc-8891-0af081417bca-25863903

CITI PROGRAM

Completion Date 08-Nov-2021
Expiration Date 07-Nov-2024
Record ID 25863898

This is to certify that:

Chang Ren

Not valid for renewal of certification through CME.

Has completed the following CITI Program course:

IRB Additional Modules
(Curriculum Group)
Research in Public Elementary and Secondary Schools - SBE
(Course Learner Group)
1 - Basic Course
(Stage)

Under requirements set by:

Auburn University

CITI
Collaborative Institutional Training Initiative
101 NE 3rd Avenue, Suite 320
Fort Lauderdale, FL 33301 US
www.citiprogram.org

Verify at www.citiprogram.org/verify/?w5e9636d9-6f71-48d1-8334-da8d8e91d5ea-25863898

CITI PROGRAM

Completion Date 15-May-2023
Expiration Date 15-May-2026
Record ID 55875700

This is to certify that:

Chang Ren

Not valid for renewal of certification through CME.

Has completed the following CITI Program course:

IRB Additional Modules
(Curriculum Group)
Internet Research - SBE
(Course Learner Group)
1 - Basic Course
(Stage)

Under requirements set by:

Auburn University

CITI
Collaborative Institutional Training Initiative

101 NE 3rd Avenue, Suite 320
Fort Lauderdale, FL 33301 US
www.citiprogram.org

Verify at www.citiprogram.org/verify/?wffeb5bfc-0bbf-4ada-bbcf-4590d25ee412-55875700

CITI PROGRAM

Completion Date	24-Jan-2023
Expiration Date	24-Jan-2026
Record ID	50321897

This is to certify that:

Chang Ren

Not valid for renewal of certification through CME.

Has completed the following CITI Program course:

Responsible Conduct of Research
(Curriculum Group)
AU Basic RCR Training for ALL Faculty, Staff, Postdocs, and Students
(Course Learner Group)
1 - RCR
(Stage)

Under requirements set by:

Auburn University

CITI
Collaborative Institutional Training Initiative
101 NE 3rd Avenue, Suite 320
Fort Lauderdale, FL 33301 US
www.citiprogram.org

Verify at www.citiprogram.org/verify/?w9412a3d2-2018-4f26-9609-01c0babda158-50321897

CITI PROGRAM

Completion Date 18-Jan-2022
Expiration Date 17-Jan-2025
Record ID 42389702

This is to certify that:

Cheryl Seals

Has completed the following CITI Program course:

Not valid for renewal of certification through CME.

IRB # 2 Social and Behavioral Emphasis - AU Personnel - Basic/Refresher
(Curriculum Group)

IRB # 2 Social and Behavioral Emphasis - AU Personnel
(Course Learner Group)

1 - Basic Course
(Stage)

Under requirements set by:

Auburn University

CITI
Collaborative Institutional Training Initiative

www.citiprogram.org/verify/?w17198472-33b9-4149-9b8d-37254afd1268-42389702

Completion Date 23-Sep-2021
Expiration Date 22-Sep-2024
Record ID 42389704

This is to certify that:

Cheryl Seals

Has completed the following CITI Program course:

IRB Additional Modules
(Curriculum Group)
History and Ethical Principles - SBE
(Course Learner Group)
1 - Basic Course
(Stage)

Not valid for renewal of certification through CME.

Under requirements set by:

Auburn University

CITI
Collaborative Institutional Training Initiative

www.citiprogram.org/verify/?wcabe5863-d5b4-4f39-a42b-5b18db20c4d4-42389704

CITI PROGRAM

Completion Date 27-Sep-2021
Expiration Date 26-Sep-2024
Record ID 42389703

This is to certify that:

Cheryl Seals

Has completed the following CITI Program course:

Not valid for renewal of certification through CME.

IRB Additional Modules
(Curriculum Group)
Conflicts of Interest in Research Involving Human Subjects
(Course Learner Group)
1 - Basic Course
(Stage)

Under requirements set by:

Auburn University

CITI
Collaborative Institutional Training Initiative

www.citiprogram.org/verify/?w80c2506c-69bc-496a-9e12-9fdaf9438f1e-42389703

CITI PROGRAM

Completion Date	16-Jan-2022
Expiration Date	15-Jan-2027
Record ID	28084134

This is to certify that:

Cheryl Seals

Has completed the following CITI Program course:

Not valid for renewal of certification through CME.

Responsible Conduct of Research for Social and Behavioral
(Curriculum Group)
Social, Behavioral and Education Sciences RCR
(Course Learner Group)
2 - RCR Refresher
(Stage)

Under requirements set by:

Auburn University

CITI
Collaborative Institutional Training Initiative

www.citiprogram.org/verify/?w4491791b-7b1c-42c2-8904-14b05eb9bce3-28084134