ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION

Paper ID #36572

Designing an ASR-based Interactive Game for Enhancing Speech Therapy to encourage young children to adhere to therapy protocols: A Case Study in User Interface Design

Chang Ren, Auburn University

Chang Ren is a lab member of the Human-Computer Interaction Lab of Dr. Cheryl Seals from the Department of Computer Science and Software Engineering at Auburn University. Chang received a master's degree in computer science from Auburn University in 2018, and currently studying for a Ph.D. Chang's current research focuses on designing an ASR-based speech training system for young children with speech disorders by incorporating gamification techniques to increase children's motivation for speech therapy through a more interactive experience.

Dr. Cheryl Seals, Auburn University

Dr. Cheryl Denise Seals is a professor in Auburn University's Department of Computer Science and Software Engineering. She graduated with a B.S. C.S. from Grambling State University, M.S. C.S. from North Carolina A&T State University, and a Ph.D. C.S. from Virginia Tech. Seals conducts research in Human-Centered Articffidial Intelligence U & HCI with an emphasis on visual programming of educational simulations, user interface design & evaluation, and educational gaming technologies. Dr. Seals also works with computing outreach initiatives to improve CS education at all levels by a focused approach to increase the computing pipeline by getting students interested in STEM disciplines and future technology careers. One of these initiatives is the iAAMCS (Institute for African American Mentoring in Computing Sciences) & STARS Alliance (starsalliance.org) with programs in K-12 outreach, community service, student leadership, and computing diversity research.

Mr. Dongji Feng, Auburn University

Dongji Feng a passionate NLP Ph.D. candidate working with Dr.Santu in BDI Lab . His interesting areas are Information Retrieval(IR), Natural Language Processing (NLP) and related evaluation metrics.

Designing an ASR-based Interactive Game for Enhancing Speech Therapy to encourage young children to adhere to therapy protocols: A Case Study in User Interface Design

Chang Ren, Dongji Feng, and Cheryl D. Seals
Auburn University Auburn, AL USA

Abstract

This research discusses an application which recognizes disordered speech with machine learning techniques. The User Interface Design (UID) course focuses on the theory and practice of designing and developing interactive systems. This project inspects the current application design and will potentially redesign a gamified system that supports speech recognition of children with communication disorders. The literature discusses successful examples of speech recognition for adults, but most systems fail with youth with speech disorders. This work has created an initial prototype to recognize speech within this population. We aim to provide novel projects for UID engineering students with interactive case studies where students review, critique, redesign, analyze and develop this system to reinforce Computer Science student programming and system design skills. Based on the interactions with community partners, we will refine the design to incorporate the model to improve engagement, provide feedback for children, and begin iterative prototype development.

Keywords

User Interface Design, Gamification, Automatic speech recognition, Speech Therapy, Machine Learning

Introduction and Literature Review

Speech therapy refers to the treatment of speech disorders and communication problems. The Automatic Speech Recognition (ASR) based approach continues to attract a lot of interest in the field of speech and language therapy. Recent advances in machine learning have made ASR a powerful and viable solution for automatically transcribing speech input for therapeutic purposes based on the earlier studies[1,2].

Existing ASR-based speech therapy systems are primarily focused on providing speech therapy to adults and achieving decent performance compared to traditional methods. These include applications such as the CHASING game developed by Ganzeboom et al.[3], which is an ASR-based serious game for providing speech therapy to elderly individuals with dysarthria. The advantage of these applications is the integration of game-based mechanisms and state-of-art ASR technology to provide interactive and automatically derived feedback in real-time to patients without the intervention of speech-language pathologists (SLPs). However, these gamified systems are not specifically designed for children which means that if corrective feedback on the game is given in the absence of SLPs or parents to facilitate their interpretations, young children will not be able to understand and tend to be less motivated to continue using their speech. Another common approach can be demonstrated by Into the Forest[4] which simply displays a virtual teacher to teach children to speak specific words in a 3D environment. In general, such applications consist of an avatar controlled by the user and may require the user to wear a headset to gain a fully immersive 3D experience. But there are manufacturer warnings against the use of this 3D technology in young children because of unclear adverse effects on visuomotor functions. And due to discomforts with headsets or motion sickness, there is a portion of young children who may not be able to endure such gameplay for a long session[5,6]. By inspecting the current applications, we found that the main challenges in designing an APR-based speech therapy system are: first, how to maintain sufficient motivation for children during long practice sessions at home, and second, how to provide feedback in a way that children can easily understand without the interpretation from SLPs or parents.

These challenges provide a rich learning experience for engineering students involving not only the user interface design (UID) of a therapeutic system for young children but also cutting-edge machine learning and speech recognition techniques. UID is the process designers use to build interfaces in software or computerized devices that focus on looks or style[7]. In our course students are introduced to design thinking and based on the school of thought we must empathize with users, define our context, and design, build and test prototypes. To investigate designs, we will review design cases to discuss important design attributes to give designers in training good examples to discuss and reinforce their understanding of good design principles. For designs with children, we need to ensure that the system has clear and specific instruction at an appropriate reading and child's level of understanding and utilizing mental models and knowledge about the world to help them utilize the application[8]. As a portion of the class, we will have case studies of many applications to provide good data and feature inspection to iteratively improve the design of the system.

In our work, the proposed design is to improve user engagement through a more interactive gamified experience, and the foundation of this application is an ASR-based speech therapy system for young children in an accessible manner that mimics the knowledge of speech therapists. Research over the past few years has increasingly focused on course content, but not much attention to this specific task-oriented system. Our focus was to develop an engaging, easy-to-use system that reduces the effort for SLPs and supports interactive game experience for the treatment of speech therapy.

Methods

UID course learning model

The UID course combines the theory and practice of interface design for interactive systems, usability engineering techniques, and implementing and evaluating interfaces. The UID course delivers instruction to upper-level undergraduate and graduate students in computer science and software engineering. Initially, we begin the course with the foundations of the interaction design theory of UID, and the class culminates with UID in practice. The course exposes the students to fully elaborated case studies and uses these as practical exercises to reinforce design theory and best practices. The learning episode is an exercise in Design Thinking that begins with requirements, design, development, and testing. Following the testing phase, project teams present their findings from preliminary user evaluations that discuss user satisfaction and system effectiveness. For the case study presented in this paper, the course teams begin their effort by reviewing the current design prototype and existing speech systems. Teams will review requirements and provide feedback for the participatory design partners (i.e., experts in communications disorders). Based on these interactions, we craft user scenarios and utilize UML (Unified Modeling Language) to capture a pictorial view of the system and catalog roles, actors, actions, and classes within a system. Upon completion of capturing user scenarios, creating software requirements, and identifying software language and environment, the development team will begin iteratively developing software to instantiate the system. The Development team will need to pilot test the system, and at the end of the first cycle of development, the team will need to have users validate that the system works as anticipated. Finally, near the end of this cycle, the design team and content experts verify that the planned scenario meets the specified requirements for the target users[9,10,11].

Speech Database

The ASR modules in the system have to deal with the disordered speech from children, which is notoriously harder to recognize than the standard speech. One of the barriers to developing ASR models that can handle disordered speech is the scarcity of datasets publicly available for training and testing, especially for young children. The Speech Exemplar and Evaluation Database (SEED) dataset were found to be the potential dataset for training and testing in this study. The corpus contains words and sentences used for clinical assessment and research of speech disorders, with more than 16,000 speech samples from adults and children with and without speech disorders[12].

ASR model

We propose a phone-level ASR model applied with high-resolution Mel-frequency cepstral coefficients (MFCCs) as features, and bidirectional long short-term memory as the model architecture for the system. Figure 1 displays the high-level overview of the ASR system. The first step in any ASR system is to extract features. In short, it identifies the components of the audio signal that are good for recognizing the linguistic content and discarding all the other stuff which carries information like background noise, emotion, etc. In this study, the model applied 40-dimensional MFCCs as features that carry the 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 training is done by Stochastic Gradient Descent with an initial learning rate of 0.01. 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.

System Design

To avoid users being deferred by unfamiliar serious games that they haven't even tried, we propose using popular games as core gameplay rather than building new games from scratch. Eventually, the user interface of our Word-Mine system was developed to improve engagement in speech therapy by using the premise of the classic Gold-Mine-like game. As a speech therapy system, it uses the movement of a mine cart, and the rewards obtained as abstract audiovisual feedback. The core gameplay is illustrated in Figure 2.

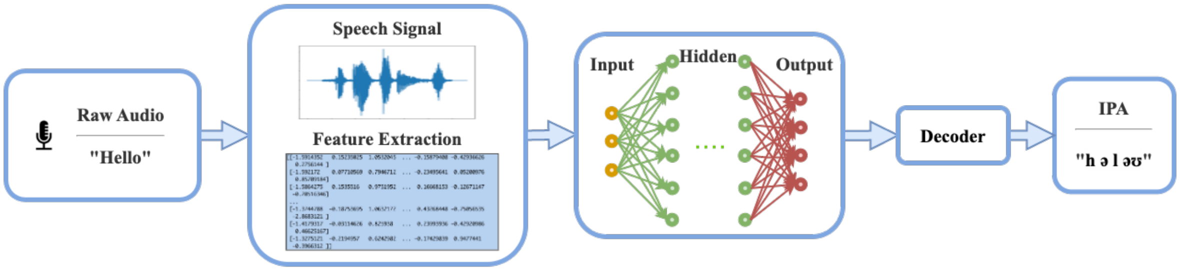

Figure 1. Overview of the phone-based ASR system

Word-Mine contains three levels representing three different difficulty levels of the therapeutic words. The words used in the system are specially selected by experts in communications disorders. After choosing a level, the system will load the vocabulary corresponding to the current difficulty. Users can listen to the standard pronunciation or record the pronunciation of the word of their choice. The APR-based recognizer is for identifying the phonetic transcription of the input speech. Every time the user completes one recording, the assessment component in the system supports a comparison between the identified phonetic symbols and the standard one to provide real-time feedback. The movement of the minecart can serve as intuitive feedback which was found easier to interpret for young children. If the child user pronounces it correctly, the minecart will move directly to the gold nugget and the child will be rewarded. Considering the frustration that young children may experience during therapy led to early discontinuation of speech therapy sessions, they could see the minecarts getting closer to the target, even though their pronunciation was problematic but better than before. The system tracks the number of improvements, and users can receive rewards after three improvements.

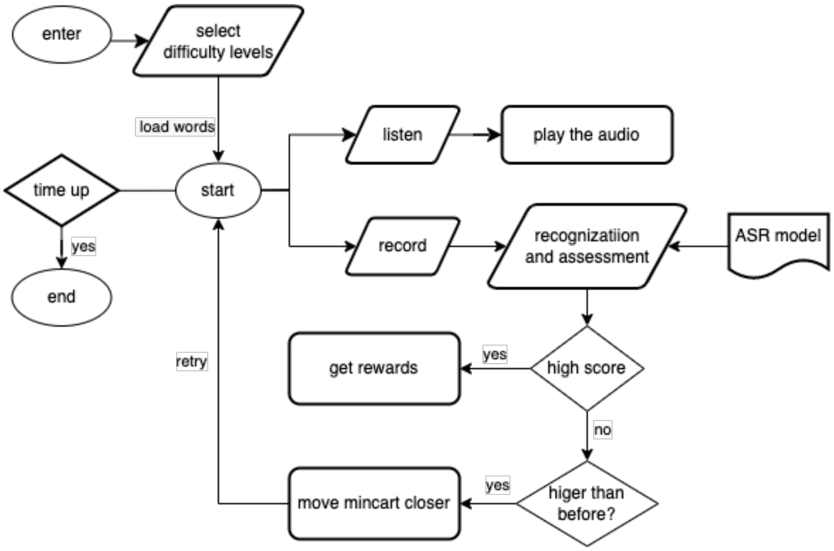

Figure 2. Flow chart of the proposed system

The initial prototype of the user interface is shown in Figures 3 and 4. The goal of Word-Mine is to collect as many gold nuggets in the gold mine as possible within the allotted time. Figure 4 shows the gameplay screen of the system. The difficulty of a word is indicated by the size of a gold nugget. To help users with standard pronunciation, they can listen to the standard speech or check its phonetic transcription by clicking the buttons in the upper left corner. Players are instantly rewarded by clearly speaking the selected word that appears on the gold nugget. In this case, the minecart goes directly to the target nugget and carries the gold for the user. In another case, the minecart will move along the track, approaching the target word if the recorded pronunciation improves. After collecting the gold, a new word will appear in the same place, replacing the old one.

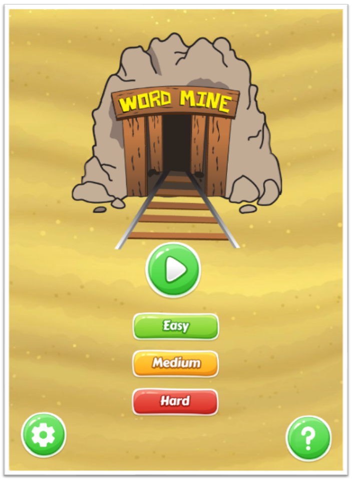

Figure 3. Main Menu

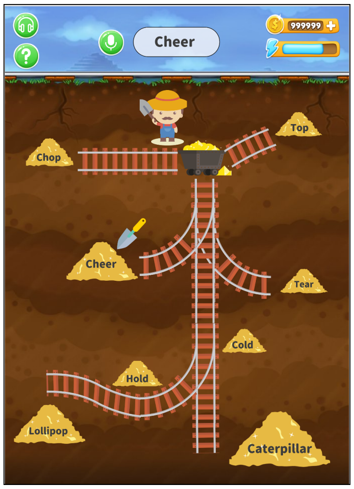

Figure 4. Gameplay Screen

Results and Analysis

We assign SEED data (i.e., 10,131 speech samples) into training and testing subsets to perform ASR experiments with a ratio of 95 and 5, respectively. With the implementation of a phone recognition for speech therapy application, our goal is not to reconstruct and correct words based on detected phones but to transcribe what the child has pronounced accurately, including potential phone-level speech errors. Therefore, instead of using the classic WER to measure performance, we use the Phone Error Rate (PER). The PER metric considers all mismatches between the recognizer hypothesis and the manual phone-level annotated reference (see definition in Equation 1), with C, I, S, and D respectively, referring to the number of correct detections, insertions, substitutions, and deletions[13].

$$\text{PER} = \frac{I + S + D}{C + S + D} \quad (1),$$

Figure 5 shows the PER, with more than 4700 child speech samples in the dataset, the PER is about 26.1%. An earlier study[13] investigated phone recognition with acoustic models of different architectures, achieving the best PER of 28.1%. Compared with this baseline, our proposed method on SEED brings an improvement of 2%.

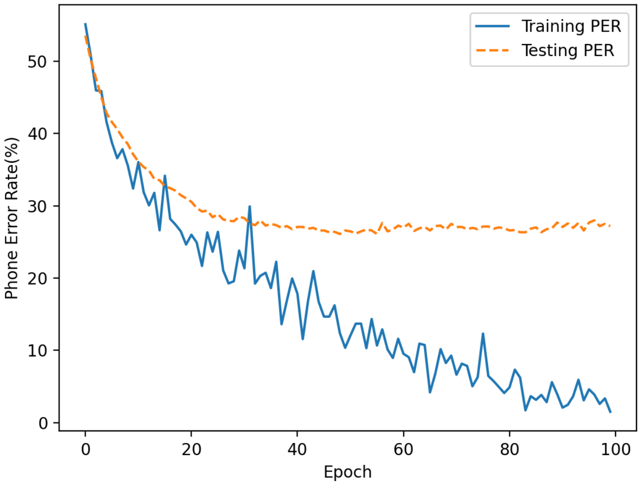

Figure 5. PER on SEED database

Concerning the system user interface, the gameplay motivates children to keep trying to pronounce words correctly, with incentives for repeating as many words as possible for more rewards. This gamification approach relies on the simplicity and appeal of the classic goldmine game to create a novel speech therapy system that increases the child's motivation to actively participate in therapy, thereby increasing the activities' effectiveness. This game-based system also allows young users to perform speech therapy in an accessible way. Moreover, the combination of cutting-edge technology and user interface design in engineering education has a very strong appeal and motivation for computer science students.

Through the study of this case in the UID course, we expose engineering students to hands-on experience in acquiring and applying industry-standard theories and methods to design successful user interfaces, experimentally evaluate the interactive system and learn the fundamentals of speech recognition technology. The UID learning model provides students with a comprehensive understanding of the user interface design process. It also provides a great experience in terms of teamwork as they have to work with teams of 4-8 people depending on the size of the project, and allows students to practice more programming, which is essential for computer science students. The additional advantage is that students improve their writing skills by creating technical reports for their projects, which is especially important for students pursuing the completion of a dissertation and is necessary for transitioning into the industry. At the same time, research experiences give students a brief introduction to the entire process of research (i.e., problem, requirements, method, analysis, presentation of solution), and students have the opportunity to review scholarly articles or conference-style papers that will inform their future writing practice[9,10,11].

Conclusions

In this paper, we have reported research aimed at developing an ASR-based interactive system that can provide speech therapy to children with speech disorders. To further improve this application, we plan to explore different ASR model architectures, such as Transformers (i.e., supporting assessment components of the system). Moreover, we will investigate the method and collect children's speech datasets in other languages, providing requirements and possibilities for the next step in making multilingual speech therapy systems.

This novel project will be introduced in the UID course and provide computer science students with practical and engaging opportunities to study, design, and develop systems with improved and gamified interfaces. Through this case study, the engineering students in the course will refine the design of the user interface to incorporate advanced interactions to provide intuitive, consistent, and automatically derived feedback, potentially increasing engagement, and intrinsic motivation for broader usage by the populations served and supporting iterative prototype development. This opportunity will strengthen students' understanding of UID principles and equip them with more UI tools and toolkits to support their career development in Software Engineering or Software Design.

References

1 Berke, Larwan, Christopher Caulfield, and Matt Huenerfauth, "Deaf and Hard-of-Hearing Perspectives on Imperfect Automatic Speech Recognition for Captioning One-on-One Meetings," In Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '17). Association for Computing Machinery, New York, NY, USA, 2017, pp. 155-164.

2 Prietch, Soraia S., Napoliana S. de Souza, and Lucia.V.L. Filgueiras, "A Speech-To-Text System's Acceptance Evaluation: Would Deaf Individuals Adopt This Technology in Their Lives?" Universal Access in Human-Computer Interaction. Design and Development Methods for Universal Access, Springer, Cham, 2014, pp. 440-449.

3 Ganzeboom, Mario, Emre Yilmaz, Catia Cucchiarini, and Helmer Strik, "An ASR-Based Interactive Game for Speech Therapy," Proceedings of 7th Workshop on Speech and Language Processing for Assistive Technologies, 2016, pp. 63-68.

4 Nasiri, Nahid, Shervin Shirmohammadi, and Ammar Rashed, "A serious game for children with speech disorders and hearing problems," 2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH), 2017, pp. 1-7.

5 Tychsen, Lawrence, and Paul Foeller, "Effects of Immersive Virtual Reality Headset Viewing on Young Children: Visuomotor Function, Postural Stability, and Motion Sickness," American Journal of Ophthalmology, 2020, pp. 151-159.

6 Roettl, Johanna, and Ralf Terlutter, "The same video game in 2D, 3D or virtual reality – How does technology impact game evaluation and brand placements?" PLOS ONE, Public Library of Science, 2018, pp. 1-24.

7 Interaction Design Foundation, "User Interface (UI) Design,", Retrieve From: https://www.interaction-design.org/literature/topics/ui-design

8 Liu, Feifei, "Designing for Kids: Cognitive Considerations,", 2018, Retrieve From: https://www.nngroup.com/articles/kids-cognition/

9 Marisha, Speights Atkins, Cheryl D. Seals, Dallin J. Bailey, "At the intersection of applied sciences: Integrated learning models in computer science and software engineering and communication disorders." Science Education and Civic Engagement: An International Journal, 2019, pp. 37-43.

10 Marisha, Speights Atkins, Cheryl D. Seals, and Dallin J. Bailey, "The Automated Phonetic Transcription Grading Tool: Where Computer Science Meets Clinical Problem Solving in Communication Disorders," SENCER Summer Institute (SSI), Santa Clara, CA, National Center for Science & Civic Engagement, 2018.

11 Marisha, Speights Atkins, Cheryl D. Seals, Dallin J. Bailey, "Fostering undergraduate and graduate interdisciplinary research in communications sciences and disorders and software engineering to develop online phonetics training modules," SENCER Summer Institute (SSI), Case Western, OH, 2019.

12 Marisha, Speights Atkins, Dallin J. Bailey, and Suzanne E. Boyce, "Speech exemplar and evaluation database (SEED) for clinical training in articulatory phonetics and speech science," Clinical linguistics & phonetics, 2020, pp. 878-886.

13 Gelin, Lucile, Morgane Daniel, Julien Pinquier and Thomas Pellegrini, "End-to-end acoustic modeling for phone recognition of young readers," Speech Communication, Elsevier, 2021, pp. 71-84.

Chang Ren

Chang Ren is a lab member of the Human-Computer Interaction Lab from the Department of Computer Science and Software Engineering at Auburn University. She received a master's degree in computer science from Auburn University in 2018, and currently studying for a Ph.D. Chang's current research focuses on approaches to design and develop an ASR-based speech training system for young children with speech disorders by incorporating gamification techniques to increase children's motivation for speech therapy through a more interactive experience.

Dongji Feng

Dongji Feng is a passionate NLP Ph.D. candidate working with Dr. Shubhra Kanti Karmaker ("Santu") in BDI Lab at Auburn University. His interesting research areas are Information Retrieval (IR), Natural Language Processing (NLP), and related evaluation metrics.

Cheryl D. Seals

Dr. Cheryl Denise Seals is a professor in Auburn University's Department of Computer Science and Software Engineering. She graduated with a B.S. C.S. from Grambling State University, M.S. C.S. from North Carolina A&T State University, and a Ph.D. C.S. from Virginia Tech. Dr. Seals conducts research in Human-Centered Artificial Intelligence & HCI with an emphasis on visual programming of educational simulations, user interface design & evaluation, and educational gaming technologies. She also works with computing outreach initiatives to improve CS education at all levels by a focused approach to increase the computing pipeline by getting students interested in STEM disciplines and future technology careers. One of these initiatives is the iAAMCS (Institute for African American Mentoring in Computing Sciences) & STARS Alliance (starsalliance.org) with programs in K-12 outreach, community service, student leadership, and computing diversity research.