Using Interactive Objects for Speech Intervention

Foad Hamidi
Department of Computer Science
and Engineering
York University, Toronto, Canada
M3J 1P3
fhamidi@cse.yorku.ca

ABSTRACT

Technological advances in physical computing and automatic speech recognition (ASR) have made the development of novel solutions for speech intervention possible. I plan to combine an ASR engine with programmable microcontrollers to develop exercises and activities based on interaction with smart objects for helping with speech therapy intervention for children.

Keywords

Physical Computing, Speech Recognition, Speech Therapy

1. THESIS RESEARCH SUMMARY

The key component of speech therapy intervention is one-on-one practice with a speech therapist. This is typically conducted in a private session where the therapist involves the client in speech activities, such as games, and provides them with feedback. The client is then encouraged to practice speech on his or her own before meeting with the therapist again. Frequent private lessons are expensive and might not be available to a lot of children. While it is impossible to replace the work of a speech therapist with a computer application, recent technological advances in automatic speech recognition (ASR) and physical computing hold great promise for the development of tools to supplement professional intervention with exercises and activities that make them more effective and efficient.

2. BACKGROUND RESEARCH

The shortage of speech training resources for people who need speech therapy is widely recognized [3]. This need is especially pressing for children since intervention works best if started at an early age. Previous research has recognized the potential of ASR for speech training [8]. Additionally, the importance of games and engaging exercises and activities for children is emphasized in previous research and used successfully in various areas of assistive technology [7, 11]. Various computer-based speech training (CBST) systems such as ARTUR [4], Box of Tricks [14] and Speech Training, Assessment and Remediation system (STAR) [6] exist that rely on ASR to implement a speech training system for children. While most of these approaches use games to engage children in speech training exercises, they still utilize the computer screen and speaker to provide feedback to the user. Furthermore, the feedback is in the form of wave-form visualizations, tongue and vocal tract animations or recorded voices. Figure 1 shows a screenshot of the feedback that is provided by the ARTUR application [4]. I believe that as the video of the speech therapist that we propose to provide as feedback is more helpful to children than these forms of feedback. In other research, the Nintendo game console is used as an alternative interactive device for training games and communication tool by children [12].

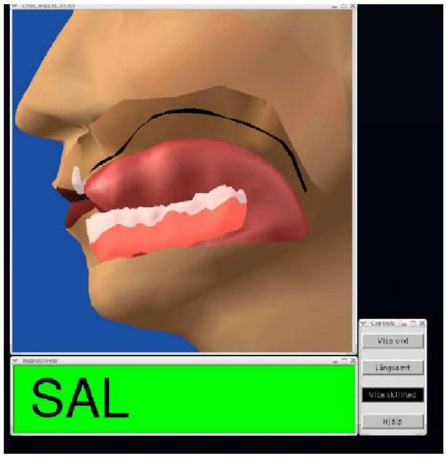

Figure 1. The feedback window of ARUTR [4]

3. SPEECH AND INTERACTIVE OBJECTS

In this research, I use tools developed in the area of physical computing to develop games that utilize novel physical interfaces, (i.e. interactive objects), that transcend the computer screen, mouse and keyboard interaction paradigm. There are several benefits for using this approach. Firstly, interacting with tangible objects emphasizes the element of play and makes the experience more immersive. Secondly, playing with these objects resembles the actual speech therapy sessions where a large number of varied toys such as dolls and drawings are used to elicit speech. Finally, these interfaces are more customizable for the specific needs of children with disabilities and their design is not confined to the limits of the desktop computer interface. For example, one possibility is to customize the object for blind users by implementing movement detection and vibration feedback in them.

For the implementation of the interactive objects, I use the Arduino microcontroller, an open-source embedded programmable controller that is used in physical interfaces [1, 10]. Recent research has demonstrated the flexibility and great potential of this controller for the development of smart textiles and interactive objects [5]. The controller is embedded in dolls and building blocks that are used in the intervention to afford speech interaction with the child. The speech is received by embedded microphones and sent over a wireless network to a computer. The actuators on the toys are activated based on the received voice; thereby implementing a dynamic interaction between the child and the toy via the computer. I use the CMU-SPHINX open-source speech recognition engine to process the input signal from the microphone [2].

An essential part of my solution is developing an interface that allows the speech therapist to configure the program to suit the need of each child. As recommended in the literature, I recognize that actively involving the speech therapist is essential for the success of the system [8]. To this end, I have developed an easy to use interface to customize the activities and feedback of the system. Figure 2 shows a screenshot of the video feedback configuration component of the interface. The therapist specifies what words or phrases are to be practiced over a practice period and records video clips demonstrating the correct pronunciation of each phrase. While the feedback provided to the child is mainly through various behaviors displayed by the interactive object (e.g., vibrations, sounds, lights), the video recorded by the therapist is also played on the computer screen if the child repeatedly fails to pronounce a word and needs help with the pronunciation.

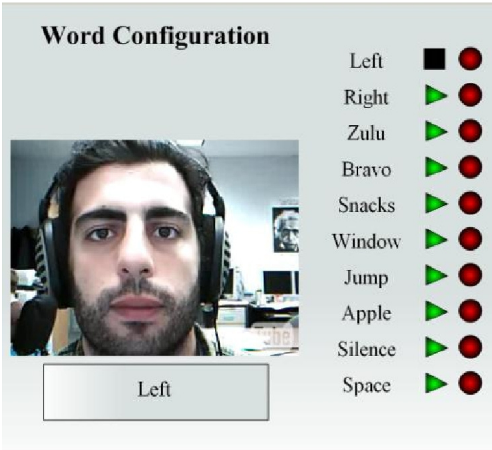

Figure 2. Video configuration interface allows the SLP to review and record video clips for each word

The speech therapist can configure the difficulty of the exercises by specifying how closely the child's pronunciations should match correct pronunciations stored in the system. The system keeps track of the interactions of the child and provides the therapist with a report at the end of the exercise period that is used to assess which tasks and games are completed and which words and sounds are problematic.

By limiting the words that are recognized by the system to a small number I simplify the recognition task and increase the accuracy of the engine as recommended in previous research [9]. This approach also allows the system to work well without training which is especially important for children and users with disabilities because successful training requires considerable effort and patience. To address concerns about the performance of the system for use by children and users with dysarthric speech I have conducted preliminary tests with both groups and have observed that the system performs satisfactorily without training. The application is designed primarily for training children but adults can also use it. The exercises can be configured to accommodate a wide range of users with various cognitive and literacy abilities.

By bringing together the fields of assistive technology and physical computing, this research paves the way for a new generation of applications that do not depend on the computer screen, mouse and keyboard as input and output devices. This approach is valuable because it can address the wide variety of the needs of people with disabilities.

I have been working with speech recognition and in particular the CMU-SPHINX engine for more than three years and have used it to implement a speech-interface for web navigation [13] So far in this project, I have created a prototype that incorporates speech recognition with the microcontroller. The next step is to develop prototypes of the interactive objects followed by user tests. The results of these tests will be used to refine the design. Finally, the refined system will be formally evaluated using a longitudinal study with actual users to assess its effectiveness.

4. ACKNOWLEDGMENTS

I would like to thank the panelists of the Eleventh International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS 2009) Doctoral Consortium, where this work was first presented, for their helpful feedback and advice. I would like to thank Dr. Melanie Baljko and Dr. Nigel Livingston for their advice and support.

5. REFERENCES

[1] Arduino. Available: www.arduino.cc.

[2] cmusphinx.org. CMU Sphinx: The Carnegie Mellon Sphinx Project. http://cmusphinx.org, Oct 2007.

[3] Editor. 2008, ASU designs program to address shortage of speech therapists, Phoenix Arizona News. DOI= http//evliving.com.

[4] Bälter, O., Engwall, O., Öster, A.M., and Kjellström, H. 2005. Wizard-of-Oz test of ARTUR - A computer-based speech training system with articulation correction. In Proceedings of the Seventh International ACM SIGACCESS Conference on Computers and Accessibility, 36-43.

[5] Buechley, L., Eisenberg, M., Catchen, J., and Crockett, A. 2008. The LilyPad Arduino: using computational textiles to investigate engagement, aesthetics, and diversity in computer science education. In Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems, 423-432.

[6] Bunnell, H. T., Yarrington, D. M., and Polikoff, J.B. 2000. STAR: Articulation training for young children. In Int. Conference on Spoken Language Processing, Vol.4, 85-88.

[7] Cockburn, J., Bartlett, M., Tanaka, J. Movellan, J. and Schultz, R. 2008. SmileMaze: A tutoring system in real-time facial expression perception and production in children with Autism Spectrum Disorder, In Proceedings from the IEEE International Conference on Automatic Face & Gesture Recognition, 978-986.

[8] Eriksson, E., Balter, O., Engwall, O., Oster, A. M., and Kjellstrom, H. 2005. Design recommendations for a computer-based speech training system based on end-user interviews. In Proceedings of the Tenth International Conference on Speech and Computers, 483–486.

[9] Franco, H., Abrash, V., Precoda, K., Bratt, H., Rao, R., and Butzberger, J. 2000. The SRI EduSpeak system: Recognition and pronunciation scoring for language learning. In Proc. of InSTIL, 123-128.

[10] O'Sullivan, D., and Igoe, T. 2004. Physical Computing. Boston, USA: Thomson Course Technology.

[11] Prensky, M. 2000. Digital game-based learning. McGraw-Hill, New York.

[12] Shirali-Shahreza, M. 2008. Aiding speech-impaired people using Nintendo DS game console. In Proceedings of the 1st international Conference on Pervasive Technologies Related to Assistive Environments, 1-3.

[13] Spalteholz, L., Li, K. F., Livingston, N., and Hamidi, F. 2008. Keysurf: a character controlled browser for people with physical disabilities. In Proceeding of the 17th international Conference on World Wide Web, 31-40.

[14] Vicsi, K., Roach, P., Öster, A.M., Kacic, Z., Barczikay, and Tantoa, A., Csatári, F., Bakcsi, Zs., Sfakianaki, A. 2000. A multilingual teaching and training system for children with speech disorders. Int. Journal of Speech technology, Vol. 3, 289-300.