Journal of Accessibility and Design for All

(CC) JACCES, 2014 – Special issue: 94-103. ISSN: 2013-7087

APPRECIATING SPEECH THROUGH GAMING

S. T. Sadural[1] and M. T. Carreon[1]

[1]Department of Computer Science, University of the Philippines, Diliman

mario.carreon@gmail.com>

***Abstract*:** This paper discusses the Speech and Phoneme Recognition as an Educational Aid for the Deaf and Hearing Impaired (SPREAD) application and the ongoing research on its deployment as a tool for motivating deaf and hearing impaired students to learn and appreciate speech. This application uses the Sphinx-4 voice recognition system to analyze the vocalization of the student and provide prompt feedback on their pronunciation. The packaging of the application as an interactive game aims to provide additional motivation for the deaf and hearing impaired student through visual motivation for them to learn and appreciate speech.

Keywords: speech recognition, hearing impairment, speech training

Introduction

Hearing impairment can happen to any child, be it by biological or circumstantial cause. Since speech is learned by children through emulating the sounds that they can hear, people tend to assume that if you cannot hear spoken language, you are unable to learn and use it.

It is a common misconception that the hearing-impaired cannot speak (Schwartz, 1987). In reality, there are some oralists in the deaf community. Through specialized teaching techniques, such as emulating how the mouth and tongue are shaped to produce certain sounds, even deaf people can learn to speak. Children with hearing impairment can, with proper training

(CC) JACCES, 2014 – Special issue: 94-103. ISSN: 2013-7087

and early intervention, overcome their difficulties, be taught and aptly trained to speak.

Most of these training techniques, however, require one-to-one interaction between teacher and student, limiting class sizes. A way around this would be through the use of a voice recognition system, in particular the SPHINX-4 (Walker et al, 2004) Hidden Markov Model speech recognition system, to listen and evaluate the speech made by children.

However, some deaf and hearing impaired students choose to stick with sign language, even if speech is being taught in their school. This is influenced by hearing-impaired individuals' belief that speech is for a 'hearing society' (Goode, 2005). The long and hard training the existing educational system uses to teach speech does not help either. This notion causes the 'deaf society's' further 'isolation' from society (Sadural, 2009).

Those who chose to break out of their stereotype are better able to integrate themselves into mainstream society better than their non-oralist counterparts do. This often results to a better lifestyle and more opportunities in the future (Sadural, 2009).

Speech appreciation, then, plays an important role in this aspect. Simply put, speech appreciation is the clear perception or recognition of the use of speech. Speech appreciation is an unacknowledged factor in a hearing-impaired student’s choice of communication mode. It directly affects the intention of the student to perform the speaking behavior (Sadural, 2009).

In light of these facts, SPREAD (Di, Gloria, Reyes, Quiniquini, 2010), or Speech Phoneme Recognition as an Educational Aid for the Deaf and hearing impaired child, is a gaming application that attempts to provide a mechanism to motivate these children to learn speech. Through a system of visual rewards, the game motivates them to try their best at their training and at the same time enjoy and appreciate speech.

(CC) JACCES, 2013 – Special issue: 94-103. ISSN: 2013-7087

Methodology

SPREAD (Di et al, 2009) is an application that, from a functional perspective, simply accepts utterances made by a user, passes the utterance to Sphinx-4 for recognition, and then displays the result. We will describe the system architecture of SPREAD in line with a typical use scenario.

The first thing a child sees when using SPREAD is the Flash-based front-end in a web browser. The child will be shown a series of simple words he or she is to pronounce. 'Apple', 'Bat', 'Star' are a sample of such words. Only one word is shown at a time. Words are also grouped into difficulty levels; higher level words can only be accessed upon completion of a lower level. Figure 1 shows the user interface for this section.

Figure 1. SPREAD main user interface

When the child is ready to pronounce the word, he or she must press the RECORD button so that the computer will record the child's attempt. Pressing the button activates the microphone as well as a Java applet which is responsible to recording the child's speech and saving it into a .wav file.

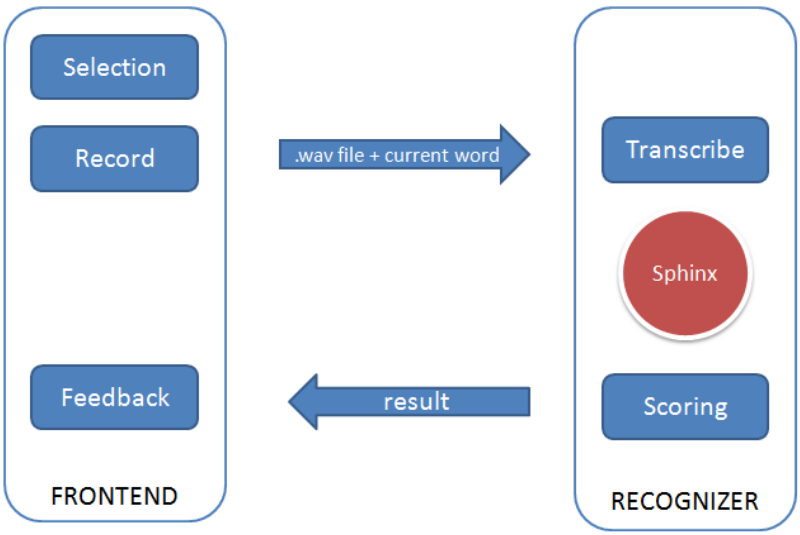
Once recording is done, the Java applet sends the recording over to the server for speech recognition. Once the .wav file arrives at the server, it is passed into the Sphinx-4 recognition engine. The server then compares the recognition result generated by Sphinx and then compares it with the

(CC) JACCES, 2014 – Special issue: 94-103. ISSN: 2013-7087

expected result. After determining the appropriate response, the server sends back the result information to the client via the Java applet. The applet then communicates with the Flash front-end to display the result.

A summary of this typical use scenario can be seen in Figure 2.

Figure 2. Functional Diagram of SPREAD



As stated earlier, SPREAD is a simple application from a functional perspective. However, with the target audience of the application being children with hearing impairment, the user experience must be designed in a way that would motivate the child to learn and appreciate speech. In the course of this work, the user interface, scoring system and the result screen are discovered to be what would make or break the speech training of the child. We will discuss these in the Results section of the paper.

Results

This section narrates the experiences gained through the exposure of SPREAD with members of the hearing impaired community. Central to this discussion is the need for an improved feedback mechanism that would effectively grade the response of the child.

(CC) JACCES, 2014 – Special issue: 94-103. ISSN: 2013-7087

Exposure to Hearing Impaired Adults

SPREAD was initially deployed for use with adult members of the deaf and hearing impaired community (Di et al, 2009), in particular, members of the Support and Empower Deaf Children, Inc. The subjects were completely deaf individuals who were not born deaf and are able to vocalize some words.

Initial feedback was very positive; the subjects were very motivated to try out the game and have expressed the wish to have had this application when they were still in school. Successful trials were even met with cheers from the subjects and their audience.

Enthusiasm to the game actually emerged as a problem as well. The excited users could not help but yell into their microphone, distorting the saved waveform data. This is in addition to the cheering audience adding noise to the data. The distorted data was difficult to be recognized by SPREAD, and even though the pronunciation was correct, SPREAD gave a negative result. Strategies to solve this can include automatically adjusting the microphone volume, giving warnings for too loud/noisy environments, or to simply to have a teacher present to guide the student into the proper use of the microphone.

This initial test also showed some urgently needed modifications to SPREAD. The first version of the application simply displayed positive/negative results (i.e. 'You got it!'/'You didn't get it…'). The negative results affected the subjects visibly, showing their embarrassment and frustration. A partial scoring mechanism was determined to be a way forward from this situation. This is discussed in a later section.

Exposure to Hearing Impaired Children

SPREAD was next deployed for use in the Special Education (SPED) division of the Batino Elementary School of Quezon City, Philippines. Unlike the adult subjects, the children were hesitant to use the software. Out of the 40 subjects, only 5 volunteered to take part once they were placed in front of

(CC) JACCES, 2014 – Special issue: 94-103. ISSN: 2013-7087

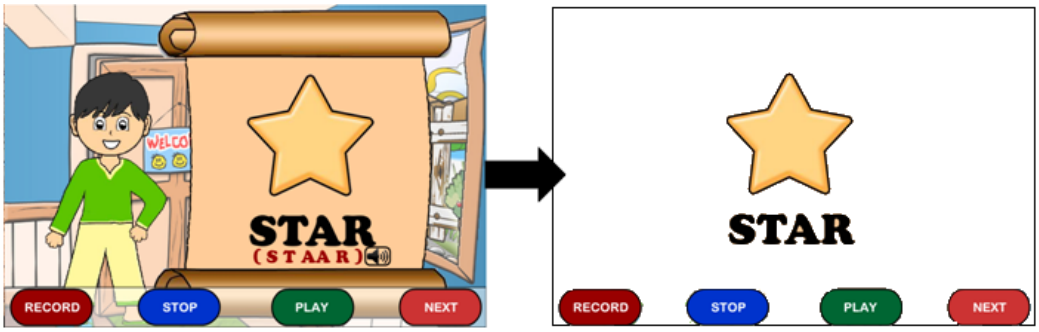
the computer. The researchers noted the shyness of most of the students; it took a little bit of coaxing to get even the five to try out the software.

It became apparent that the children did not know some of the words. Although they are able to sign most words, they were only able to speak only the very common words such as 'Car' or 'Star'. The volunteers actually treated the application as a sudden surprise test that they were not prepared for. At the end of the trials, the students were able to recite conversational messages like 'Thank you' and 'Goodbye' better than how they pronounced the words in SPREAD.

This result indicates that the students encountered in this particular test learned speech more for its utilitarian aspect rather than as for casual conversation. Simple greetings only require a limited vocabulary of spoken phrases. In casual conversations, these children opted to use signing rather than memorizing a large vocabulary of spoken words. SPREAD can be modified in the future to test out these short phrases.

There is a need too for SPREAD to be integrated with the existing speech curriculum. The students tested were unprepared for the recitation of the particular words being used by the game. SPREAD would be more effective if the words being shown to the students were just taught by their teacher or as a post lesson evaluation tool.

Figure 3. Simplified User Interface



The feedback screen also needs to be, at the least, reworded. The 'You didn't get it...' message came out as too negative for some of the users, adding to their frustration. It was recommended that instead a 'You can do

(CC) JACCES, 2014 – Special issue: 94-103. ISSN: 2013-7087

better!' message or 'Good try!' message would lessen the brunt of a wrong recording session.

On Scoring

To discuss the scoring mechanism of SPREAD, we have to first discuss how Sphinx evaluates sound data. Given the sound file, Sphinx first compares the sound with stored sound samples from the acoustic models. These stored sound samples are called phonemes, which are the basic building blocks of a spoken word. Sphinx actually produces multiple results as it tries to determine which best phoneme combination closely matches the inputted sound.

To help with the evaluation, Sphinx uses a grammar file. This grammar file tells Sphinx what particular words are expected to appear and in what particular combination. This limits the possible outcomes that come out of the decoding process. By default, Sphinx will always produce as an outcome one of the possible words in the grammar file. After Sphinx produces the best matched word, SPREAD then compares the decoded word with the expected result. If these two do not match, then SPREAD will send a negative result to the front end module.

Unfortunately, this particular scheme does not give partial points. An alternative scoring system was attempted wherein the speech was deciphered at the phoneme level instead of on a word level (Carreon, 2011). The scheme was to provide full points if all the correct phonemes of the word were pronounced correctly, partial points if the 'training phoneme' appears (in contrast to 'training word'), and a negative result if the training phoneme was not detected at all.

This scheme, however, resulted in lower recognition rates. On a per word level, Sphinx can compare the detected phonemes and match it to the closest possible word; this eliminates noise and other ambiguities as the set of possible results is only limited to a few words. The set of possible results expands exponentially on a per phoneme level as Sphinx can no longer get

(CC) JACCES, 2014 – Special issue: 94-103. ISSN: 2013-7087

any contextual clues from the grammar file and simply returns any and all phonemes it can detect.

This area is the current focus of the research. The use of the Sphinx confidence score metric seems to be a promising avenue for exploration. A simpler scheme would be to record multiple trials of the same word return how many were detected correctly vs. the total number of trials.

Conclusion

This paper discussed SPREAD and how it uses gaming as a strategy for motivating a hearing impaired child to learn how to speak. In the course of the research, it was shown that SPREAD can and does promote speech appreciation, though improvements with the user interface and feedback mechanism would have to be addressed in the near future.

Development for SPREAD is still continuing. The end goal for SPREAD is for it to be deployed as a teaching tool working in line with the traditional methodology for teaching speech. It aims only to enhance the learning experience of a child and not to supplant nor to totally replace the need for the standard speech training being done in oralist schools. SPREAD can also be made as a platform for other types of gaming strategies above and beyond this simple object identification scheme. It is hoped that this work will inspire others to explore this promising field of research.

Acknowledgments

This work has been assisted by the Philippine Department of Science and Technology (DOST) Engineering Research and Development for Technology (ERDT) Faculty Research Grant.

(CC) JACCES, 2014 – Special issue: 94-103. ISSN: 2013-7087

References

[1] Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J Kuhl & J Beckman (Eds.), Action-control: From Cognition to Behavior (pp. 11-39). Heidelberg, Germany: Springer.

[2] Ajzen, I. (2002). Perceived Behavioral Control, Self-efficacy, Locus of Control, and the Theory of Planned Behavior. Journal of Applied Social Psychology, 32, 665-683.

[3] Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. In D. Albarracín, B. Johnson & M. Zanna (Eds.), The handbook of attitudes (pp.173-221). Mahwah, NJ: Erlbaum.

[4] Bloodstein, O. (1979). Speech Pathology: An Introduction. Boston, USA: Houghton Mifflin School.

[5] Carreon, M. (2011). Alternative Scoring Systems for SPREAD Application. Proceedings of the 2011 UP College of Engineering Professorial Chair Colloquium, University of the Philippines, Diliman.

[6] Di, R., Gloria, R., Reyes, R., & Quiniquini, M. (2010). Speech and Phoneme Recognition as Educational Aid for the Deaf and Hearing Impaired (SPREAD). Unpublished undergraduate thesis, University of the Philippines, Diliman.

[7] Goode, E. (2005). Deviant Behavior (7th ed.). NJ, USA: Prentice Hall.

[8] Levy, J. (1998). Family Response and Adaptation to a Handicap. In J. Gerring & L. McCarthy (Eds.), The Psychiatry of Handicapped Children and Adolescents: Managing Emotional And Behavioral Problems (pp. 224-227). Massachusetts, USA: College-Hill Press.

[9] Mashie, J., Vari-Alquist, D., Waddy-Smith, B., & Bernstein, L. (1998). Speech Training Aids for Hearing-impaired Individuals: III. Preliminary Obesrvations in the Clinic and Childrens' Homes. Journal of Rehabilitation Research, 25(4), 69-82.

[10] Sadural, S. (2009). Measuring Speech Appreciation of Hearing-Impaired Students in Baguio School for the Deaf: A Quantitative Study. Unpublished undergraduate thesis, University of the Philippines, Baguio.

[11] Silverman, F.H. (1995). Speech Language and Hearing Disorders. USA: Allyn and Bacon.

(CC) JACCES, 2014 – Special issue: 94-103. ISSN: 2013-7087

[12] Schwartz, S. (1987). Choices in Deafness: A Parents' Guide. USA: Woodbine House.

[13] Walker, W., Lamere, P., Kwok, P., Raj, B., Singh, R., Gouvea, P., Wolf, P., et al. (2004), Sphinx-4: A flexible open source framework for speech recognition. (Sun Microsystems Technical Report, TR-2004-139).