A Speech Training Software Aiding the Hearing-impaired in Learning Vowel Sounds

A Thesis Proposal
Presented to
the Faculty of the College of Computer Studies
De La Salle University Manila

In Partial Fulfillment
of the Requirements for the Degree of
Bachelor of Science in Computer Science

by

CABERIO, Aldrin
CRUZ, Lelis
FLORES, Joshua
LIM, David

Solomon SEE
Adviser

April 19, 2015

Abstract

The hearing-impaired are not familiar with the way words are spoken because they do not know how different words sound like. One way to help the hearing-impaired is through speech training systems which teaches them how to speak. The proponents will create a speech training software that provides visual scalar feedback in order for the hearing-impaired to adjust to speaking different sounds. The speech training software will be limited to supporting only vowel sounds.

Keywords: visual feedback, speech training system, hearing-impairment

Chapter 1

Research Description

This chapter discusses the overview of the current state of technology, research objectives, scope and limitations of the research, and significance of the research.

1.1 Overview of the Current State of Technology

Hearing impairment is the inability to hear, and it may affect ones development when learning how to speak (Lasak, Allen, McVay, & Lewis, 2014). Some people are born with hearing impairment, and some obtain it due to age, illness, trauma, and other factors. An extremely deaf child may find it immensely difficult to learn speech and language since they cannot imitate any sounds without auditory feedback (Bernstein, Goldstein, & Mahshie, 1988). This leads to children not being able to use their vocal chords even though it is properly functioning.

Early speech training software were implemented using a lip reading technique, wherein the systems take the users lip movement as input (Heracleous, Beautemps, & Aboutabit, 2010). However, many phonemes share similar facial and lip shape (visemes) with other phonemes as well; resulting to phonemes not being discernable through visual data alone (Heracleous et al., 2010).

Bernstein et al. (1988) mentions that visual information is commonly practiced to aid in speech perception. Several visual displays were designed to grasp the attention of children wherein they play games to learn. One of the games is a ball game wherein the ball changes in size depending on the loudness of the sound the child makes. The ball also can be controlled going to the hoop using the childs voice pitch (Bernstein et al., 1988). Another game is a vertical spectrum, which is visually displayed as a changing 2-dimensional shape wherein frequency was displayed on the y-axis, and amplitude—the degree of change—on the x-axis (Bernstein et al., 1988). The problem with this game is that it only leaves the children to guess on whether he is actually doing right or wrong. There is no specific feedback on how to produce the right sound correctly.

Currently, one way for a hearing-impaired person to learn how to speak is by planting an electronic device called a cochlear implant. It is designed to function as a cochlea, a part of the ear which receives sounds in the form of vibrations and these vibrations are sensed by hair cells. Deaf people are unable to hear because their hair cells cannot detect the vibrations as received by the cochlea due to damaged hairs cells, or they have less hair cells. Cochlear implants stimulate the cochlear nerves without going the hair cells anymore. A person with a cochlear implant may still undergo training and therapy in order to adjust to their hearing (Blume, 2009). Once a person hears properly, it can lead to speech training and therapy in order for one to respond accordingly to different sounds.

Computer-based speech therapy systems that have been implemented are mostly product-oriented, providing an analysis of the input of the user and the final result. The method, however, does not provide the information of how the sound should be articulated. The shortcoming of these systems are brought upon by giving visual representations in real time i.e. game-like visualizations and speech pictures (Oster, 2006).

Moreover, in training hearing-impaired children, the most preferred feedback is through visual modality. This allows the children see the motor gestures that were non-visible, assisting them in developing their speech gestures. The problem that this method posed was that it was seldom evaluated in a pedagogical programme. Often, the visual aids presented was difficult to understand, unnatural, delayed, unattractive, and had no motivational impact on the children (Oster, 2006).

1.2 Research Objectives

1.2.1 General Objective

To implement a feedback system that provides scalar visual feedback by means of colour representation.

1.2.2 Specific Objectives

1. To find schools for the hearing-impaired who may assist the group with the testing and implementation of the research.
2. To identify and compare different methods and design of visual feedback in speech training systems.
3. To select the optimal colour representation to be used as visual scalar feedback.
4. To select the optimal visual feedback method based on usability and effectiveness
5. To evaluate the accuracy of the implemented speech training softwares visual feedback.
6. To evaluate the effectiveness of the implemented speech training software

1.3 Scope and Limitations of the Research

For the implementation of the speech training software, the master's thesis of Natalie Agustin (2014) will be used. Agustin's program is a mapping software which receives voice input from the user and identifies the vowel sound made by the user and displays the users articulation—the movement of the lips, tongue, jaw, and other speech organs. Agustin's thesis will be used for the system to know which vowel the user had spoken. The speech training software will be written in Java, a cross-platform programming language, making the system usable by different operating systems. The speech training software will run on a desktop platform, as desktop computers are accessible and widely-used. The proponents will review only speech training software with free trials. The feedback system will only cater to vowel sounds, as the mapping software to be used only serves vowel sounds.

1.4 Significance of the Research

The thesis will assist the Deaf in learning the pronunciations of vowel sounds correctly. Deaf communities may make use of the software, guided by a professional speech trainer, in his or her development with regard to speech training. The thesis may serve as a basis for future implementations of Deaf speech training software which may cater not only vowels, but also consonants.

The thesis will assist the Deaf in learning the pronunciations of vowel sounds accurately. Current speech training software in the industry only shows a graphical illustration of the user's voice input and the goal the user must achieve. This method however lacks in providing scalar feedback of the input of the user. The thesis will be providing a feedback on what sound is the user currently producing and what he/she needs to do in order to correct himself/herself. Deaf communities may make use of the software, guided by a professional speech trainer, in his or her development with regard to speech training. The thesis may serve as a basis for future implementations of Deaf speech training software which may cater not only vowels, but also consonants.

Chapter 2

Review of Related Literature

This chapter discusses studies for an effective speech training software and different software that have been distributed and used by the hearing-impaired in order to learn how to speak.

2.1 Review of Related Paper

2.1.1 Speech Training Systems

Speech is not innate for children who are deaf or with a profound hearing loss at birth. The effect of such disability obstructs a child from imitating other peoples sounds and comparing the sound he/she may be able to produce with theirs. However, the development which cant be learned through spontaneous speech could be acquired through vision, tactile sensation, and residual hearing. Although, this learning method solely relies on the visual representation of phonemes and through tongue control movements in sustaining the speech movements. Computer-based speech therapy aided programs makes use of visual feedbacks to allow individuals with profound hearing impairments to evaluate themselves with their produced input compared with an acceptable input. By garnering the attention of the children through drawings and illustrations to demonstrate loudness, pitch contour, spectral distribution, etc., they are able to validate their produced sound (Oster, 2006).

Speech Training Systems were developed in order to represent feedback to people with profound hearing impairment in a visually appealing manner (Oster, 2006). Other implementations of software address this solution by introducing gamifying elements to allow the software to be used by the direct users or under the assistance of teachers. Methods such as spectrographs, dating since 1947, have also been used to teach speech to children (Javkin, Antonanzas-Barroso, Das, & Zerkle, 1993). On the other hand, tactile aid focuses on somatosensation; by using vibrators at various parts of the body to indicate numerous elements of speech such as voicing or nasalization (Wankhede, 2014).

Strategies are often required in implementing such systems, especially to Deaf children, who necessitates more efficient methods of instructions compared to hearing children. Speech training is efficient if it would be able to allow children to imitate invisible speech articulations, which could not be perceived properly by visual aids (Oster, 1996). Oster (1996) suggests that for a speech training system to considered efficient and to able to amplify the any possibility for a child to learn, a number of requirements must be achieved:

- Clear instructions and pedagogical manuals must be created and made available for use with different groups of children.
- The visual feedback of the childs voice and articulation should be shown immediately and without delay.
- The aid must be acceptable to the therapist as well as to the child, which means that the aid must be attractive, interesting, easily comprehensible, easy to handle, and motivating.
- The visual pattern must be natural, logical and easily understandable. This means that training parameters as, e.g., pitch could be shown vertically as pitch variations occur; intensity through the size of an object that becomes larger as a sound becomes louder and smaller as a sound becomes softer; intonation and stress through a continuous red curve; duration could be shown horizontally and voicing through a relationship between voicing and the change of a colour.
- The aid should provide a contrastive training, that is, the correct model of the therapist and the deviant production of the child are shown simultaneously and compared with each other.
- The aid should provide a flexible, individual, and structural speech and voice training and give an objective evaluation of the childs training results.

A speech training software that provides visual aid will help a hearing-impaired person evaluate and correct his utterance or pronunciation (Wankhede, 2014).

Wankhede mentions that how the feedback is presented also affects how hearing-impaired children may improve their pronunciations. The visual feedback must also be shown immediately on the computer screen without delay to prevent confusion to the child. Some children may find some speech training aids as difficult to understand, unnatural, delayed, unattractive, and unmotivating to the children. For evaluation of the speech training aid, it should have the acceptance of a speech therapist and as well as of a child, meaning that the system should be appealing, presentable, easy to use, and motivating (Wankhede, 2014).

Aside from visual feedback that are used as aid in speech training software, there is also another type of feedback being used as aid for the hearing-impaired people in speech training - vibrotactile feedback. The Haptic Chair is a project developed by Suranga Nanayakkara, Lonce Wyse, and Elizabeth A. Taylor (2012), to help deaf people in speech training by the use of sending vibrotactile feedback to several parts of the body such as palm and fingers.

2.1.2 Self-Organizing Maps (SOMs)

Self-organizing map (SOM) is an artificial neural network that discovers patterns in the input data and can learn even without supervision (Kohonen, 1988). The input is then transformed into a one or two dimensional map and a weight is given to it and perform this transformation adaptively in a topologically ordered fashion (Chandar, Suriyanarayanan, & Manikandan, 2013).

SOMs have neurons or nodes in which each node/neuron has a weight that is similar in dimension as the input and these nodes are then arranged in the map to from a hexagonal or rectangular grid, connecting them to each input node (Agustin, 2014).

At the start, the neurons are given random weight values. This is called the initialization phase. The next step of the process is determining which neuron is closest to the input via its given weight value. "The weights of the winning neuron and neurons close to it in the SOM lattice are adjusted towards the input vector" (Chandar et al., 2013). This is called the competition phase (Agustin, 2014). Lastly, the weight of the winning neurons and its neighbors are adjusted in relation to the input pattern in this third phase called adaptation phase, hence, the SOM is created.

2.2 Review of Related Software

2.2.1 SpeechViewer

SpeechViewer III is a software developed by IBM to help the hearing-impaired enhance their vocal skills. It transforms spoken words and sounds into imaginative graphics (Speechville, 2014). "SpeechViewer III uses visual and auditory feedback to analyse and improve the speech skills of people who have speech, language or hearing disorders" (Kennedy, 1998). The student must be facilitated by a professional in guiding him or her in speech training. The software provides a dozen exercises for the user and also provides immediate and clear feedback. It also gives interest for children as the software gives animated rewards during successful responses.

2.2.2 Speech Therapy 5

"Speech Therapy uses over 70 voice-activated video games to provide real-time reinforcement of a client's attempts to produce changes in pitch, loudness, voiced and unvoiced phonation, voicing onset, maximum phonation time, sound and vowel tracking"(Tiger DRS, 1998). The software is aimed for the development of children, wherein they play interactive games. Immediate animated feedback is displayed to show how well a child performs.

2.2.3 Video Voice

Video Voice is a speech training software developed by Micro Video Corporation. It is a speech development tool which offers a wide variety of entertaining and motivating games in order for the student to learn. It provides visual feedback on pitch, volume, and vowel production.

2.2.4 Summary Table

Table 2.1 provides a summary of speech training software that were reviewed.

Table 2.1: Table of Comparison

Name of Software	Features	Means of Visual Feedback
SpeechViewer	Provides a dozen exercises for the user; clear and immediate feedback; provides animated rewards during successful responses	Visual and auditory feedback
Speech Therapy 5	70 voice-activated games; real-time feedback; aids in pitch, loudness, voiced and unvoiced phonation; voicing onset; maximum phonation time; sound and vowel tracking	Immediate animated feedback
Video Voice	Wide variety of games; immediate feedback; aids in volume, pitch, and vowel production	Immediate visual feedback

Chapter 3

Research Methodology

This chapter discusses the methodology to be followed for the entire duration of this research.

3.1 Research Activities

The research activities will be divided into the following phases:

1. Gathering of Data
2. Research and Comparison
3. Implementation of the Software
4. Testing and Evaluation of the Software
5. Analyzing the Effectivity and Usefulness of the Software

3.1.1 Gathering of Data

The appropriate literatures—scalar feedback, visual appeal, speech training systems—are to be gathered in order to design and implement an effective speech training software. Several textbooks and journals are available in the institution where the group is studying in. The group also has to look for Deaf schools who are willing to let the researchers conduct interviews with their students for the sole purpose of creating an effective speech training software.

3.1.2 Research and Comparison

After having gathered data, the group is then tasked to research and compare different literatures and ideas. There are several ways of providing visual and scalar feedback to users. However the group should only select the more effective ways of providing visual and scalar feedback. With this, it is important to look behind the science of these visual feedback and how effective they are.

There are several speech training software created by different companies. IBM, for one, created SpeechViewer which is used by speech therapists to train the hearing-impaired. The software provides a feature wherein the therapist can record training sessions in order to see the improvements of his or her patients. Though the software may be outdated, it may serve as a reference on how a speech training process works. Motivation is very important in speech training. If the software appears to be dull and boring, the user may have the lack of will to learn and develop. The proposed speech training software the group will implement must be educational, motivating, and effective. Viewing other speech training software may help the thesis group gather ideas in order to create such that.

3.1.3 Implementation of the Software

The speech training software will be implemented such that it should be appealing and functional. Also, the software must be effective and easy to use. The mapping software of Natalie Agustin (2014) will be integrated into the speech training software in order to recognize voice input and provide the corresponding mapping which determines the vowel sound that was spoken by the user.

3.1.4 Testing and Evaluation of the Software

To determine how effective the implemented software is, the researchers are going to have the program tested by the hearing-impaired. The researchers would like to know if the software is good in terms of usability, effectivity, design, and motivating. This way, the researchers will know how to further improve the software.

3.1.5 Analysis of Results

The researchers will analyze how helpful and effective the speech training software is through surveys and interviews with their testers. As mentioned in the previous subsection, the researchers will ask how usable and effective the program is and if the program is appealing and motivating.

3.2 Calendar of Activities

Table 3.1 shows a Gantt chart of the activities. Each bullet represents approximately one week worth of activity.

Table 3.1: Timetable of Activities

Activities (2015)	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Gathering of Data	●●●●	●●	●●●							
Research and Comparison	●●●●	●●	●●●							
Implementation of the Software			●●	●●●●	●●●●	●●●●	●●●●	●●●●	●●●●	●
Documentation	●●●●	●●	●●●	●●●●	●●●●	●●●●	●●●●	●●●●	●●●●	●

Table 3.2: Timetable of Activities

Activities (2016)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug
Testing and Evaluation of the Software	●●	●●●●	●●●	●●				
Analysis of Results					●●	●●●●	●●●●	●●
Documentation	●●	●●●●	●●●	●●	●●	●●●●	●●●●	●●

References

Agustin, N. (2014). *Using self-organizing maps and regression to solve the acoustic-to-articulatory inversion as input to a visual articulatory system* (Unpublished master's thesis). De La Salle University.

Bernstein, L. E., Goldstein, M. H., & Mahshie, J. J. (1988). Speech training aids for hearing-impaired individuals. *Journal of Rehabilitation Research and Development*, *25*(4), 53–62.

Blume, S. S. (2009). *Artificial ear : Cochlear implants and the culture of deafness.* Rutgers University Press.

Chandar, S., Suriyanarayanan, A., & Manikandan, M. (2013). Speech recognition using som and actuation via network in matlab. *International Journal of Electronic Signals and Systems*, *3*.

Heracleous, P., Beautemps, D., & Aboutabit, N. (2010). Cued speech automatic recognition in normal-hearing and deaf subjects. *Speech Communication*, *52*(6), 504–512.

Javkin, H., Antonanzas-Barroso, N., Das, A., & Zerkle, D. (1993). A motivation-sustatining articulatory/acoustic speech training system for profoundly deaf children. *IEEE*, 145–148.

Kennedy, G. (1998). *Speech viewer 3 review.* `http://www.spectronics.com.au/article/3073`. (Accessed March 16, 2015)

Kohonen, T. (1988). The 'neural' phonetic typewriter. *Computer*.

Lasak, J., Allen, P., McVay, T., & Lewis, D. (2014). Hearing loss: Diagnosis and management. *Primary Care*, *41*(1), 19–31.

Nanayakkara, S., Wyse, L., & Taylor, E. (2012). The haptic chair speech training aid for the deaf. *OZCHI*. `http://www.sutd.edu.sg/cmsresource/idc/papers/2012_The_Haptic_Chair_as_a_Speech_Training_Aid_for_the_Deaf.pdf`.

Oster, A. (1996). Clinical applications of computer-based speech training for children with hearing-impairment. In *Proceedings of icslp-96, 4th int. conference on spoken language processing* (pp. 157–160).

Oster, A. (2006). *Computer-based speech therapy using visual feedback with focus on children with profound hearing impairments* (Unpublished doctoral

dissertation). KTH School of Computer Science and Communciation.
Speechville. (2014). *Software: for speech-language pathology.* http://www.speechville.com/speech-therapy/software-slps.html. (Accessed March 16, 2015)
Tiger DRS, I. (1998). *Speech therapy 5.* http://www.drspeech.com/SpeechTherapy5.html. (Accessed March 16, 2015)
Wankhede, N. (2014). Designing visual speech training aids for hearing impaired children. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, *3*(4).