# Thai Language E-Training for the Hard of Hearing

Nalinrat Witsawakiti, Atiwong Suchato and Proadpran Punyabukkana

Spoken Language Systems Research Group
Department of Computer Engineering, Chulalongkorn University
Phyathai Road, Pathumwan, 10330, Bangkok, Thailand
Nalinrat.W@student.chula.ac.th, Atiwong.S@chula.ac.th, Proadpran.P@chula.ac.th

***Abstract***

*People who are hard of hearing usually have some amount of difficulty in uttering words due to the lack of auditory feedback. Speech training is needed for them to perform reasonably effective verbal communication. To learn how to utter a sound, most hard-of-hearing learners rely on lip-reading in order to imitate how the sound is produced. Traditionally, such training needs to be done on a one-to-one, face-to-face basis, especially for children. There are some computer programs facilitating this type of training. However, they are generally not well-matched with the phonemes of Thai language. This paper aims at describing the concept of an e-training tool that helps hard-of-hearing children learn and practice to utter simple Thai words more correctly. The concept of this e-training tool emphasizes on the use of technologies, such as speech recognition and computer animation, to eliminate limitations of the traditional therapist-based speech training on three aspects: Teaching, Training, and Tracking.*

## 1. Introduction

Hard-of-hearing people can still hear sounds although they face some amount of hearing loss. Thus, unlike people with normal hearing, they usually have some difficulties uttering words correctly. This is simply due to their defective natural auditory feedbacks. The characteristics of hearing loss vary from people to people. Hearing aids might be used to compensate for such defective auditory systems. Assisted with hearing aids, hard-of-hearing people are generally able to hear the sound with better clarity and more intelligibility. However, it is quite clear that for hard-of-hearing children, hearing aids do not offer them sounds with good enough resolutions that are necessary for them to reproduce the sounds verbally. This is one of many obstacles for them in language learning and, in turn, effective verbal communication with general society.

In schools, children who are hard of hearing may adopt a sign language and/or an oral language as their means of communication. Learning to communicate via a sign language is obviously an easier choice. However, opting to do so also has some equally obvious downfalls. First of all, most people with normal hearing do not know how to 'speak' sign languages. Together with the fact that once a child starts to adopt a sign language, he or she is likely to shy away from learning oral languages since the latter choice is more difficult to learn, children who only learn how to communicate with gestures are likely to be isolated from the general society. By also choosing the more difficult choice in adopting oral languages, hard-of-hearing children would be able to live their lives more like normal people. It is also important to note that they are capable of learning oral

language as they are as smart as others but special kind of teaching technique has to be developed in order to best facilitate their deficiency. With limited resource of specially trained teachers, a limited number of students can be served at a time as this kind of training needs to be performed on a one-to-one basis.

Putting limited personnel resources in the form of speech therapy specialists aside, teaching hard-of-hearing children to utter sounds of a language correctly still faces some other difficulties, such as how to clearly visualize the associated speech production mechanism so that learners can understand what is going on inside their vocal tract and how to correctly use their vocal apparatus to produce such sounds. The mentioned issue is not helped by learners just lip-read the teachers.

With advancements in some researches in key computer techniques including speech analysis, voice recognition, as well as computer graphics and animations, together with the growing of information and communication technology (ICT) usage in Thailand, it is not much less than ideal to enhance traditional speech training procedure of the hard of hearing with some computer-based tools. Therefore, this paper aims at proposing some key concepts in developing such an e-training tool. We will discuss requirements and possible solutions in three aspects: teaching, training, and tracking. Furthermore, we will present our current progress in the research and the implementation of our one-of-a-kind e-training tool at the department of computer engineering, Chulalongkorn University.

## 2. Background

Human speech production mechanism relies on the placement as well as the movement of many organs in the vocal tract. These speech production-related organs are called articulators. Major articulators associated with the production of Thai speech include the tongue, the soft palate, the hard palate, the jaw, the teeth and the lips. These articulators are placed and moved according to the sounds the speaker wants to produce. The position of these articulators manipulates the airflow supplied from the lungs so that the manipulated airflow sounds as desired when coming out of the lip opening. Figure 1 shows MRI images of vocal tract shape when two vowels are produced. We can notice the difference in the height of the tongue placement between the two vowels. This is an example showing that when uttering different sounds, a speaker places their articulators differently. On the other hand, when intending to utter a specific sequence of sounds, articulators needs to be placed in the right locations. Apparently, to learn how to pronounce specific sounds, learners need to be aware of the placement of each related articulator.

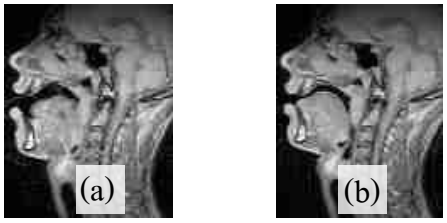


**Figure 1**[1]**:** MRI images of the vocal tract when the vowels (a) /a/ and (b) /i/ are produced

## 3. Related Work

Apart from Thai language, efforts have been spent in research and development of computer programs aimed at helping hard-of-hearing individuals learn how to utter sounds in different languages. Soleymani et al. [1] built a trainer of English vowels for hearing-impaired children, called Speech Illumina Mentor (SIM). This program

[1] From MRI gallery excerpt of Engwall, O.(http://www.speech.kth.se/~olov/mri.html) retrieved July 7th, 2006

presents the English vowel speech training in the form of computer games. English vowel learning is also a focus of another computer program called BALDI [2], in which a 3D model of a human head is used to show appropriate locations of articulators associated with each vowel to be uttered. A tool called ICATIANI applied the BALDI engine to the training of five Mexican Spanish vowels [3]. A speech training aid utilizing 3D computer animation called ARTUR was also developed for Swedish [4]. Jiang et al. [5] developed a program called STODE (Speech Training for Oral Deaf Education) which aims at reducing the intonation confusion and pronunciation timing errors in Mandarin. Another computer program that assists the training of Mandarin for the hearing impaired in home-school environments was developed by Hsiao et al. [6]. This program also keeps the logging for each learner's progress.

## 4. Motivations

### Language-specific tools

Even though there are some tools that help the hard of hearing practice articulating their speech, they are developed based on foreign languages. Many schools in Thailand with hard-of-hearing students tried such foreign-made programs with unsatisfactory results. The reason lies in the nature of Thai and foreign speech. The acoustic models which are required for such tools to automatically analyze the learners' speech are built (or, in a technical speech recognition jargon, trained) upon foreign languages. Although, some languages, including English, share many phonemes with Thai, the acoustic properties of those foreign phonemes are, in fact, different from the ones of Thai.

### Why E-Training?

As mentioned earlier, one of the problems in the speech training for the hard of hearing is the limited specialist personnel resources. This results in a lack of continuity in the learning and practicing process. Therefore, electronic-based training is an ideal choice for the task due to a similar reason that makes eLearning very popular. The reason is that it provides learners with opportunity to learn and/or practice whenever and wherever they would like. This reason is even more favorable in the case of the hard of hearing due to their special needs. Furthermore, e-training can be utilized with limited supervision. It is possible that parents of the hard of hearing can supervise their children's learning and practicing in their own home, instead of having to be assisted by specialists at school, while the specialists can still monitor their progresses via the internet.

### The need for integrated systems

Most of the speech training tools are stand-alone tools with focuses on some specific aspects of speech training. However, it is obviously a better alternative to build an integrated speech training system that combines special requirements of the hard of hearing with other traditional mainstream requirements for good eLearning tools. Advancements in eLearning methodology should be taken into consideration when building such systems. This does not only applies to the speech training systems, but also to any training systems for people with special needs.

## 5. Requirements and Solutions

Such benefits of using electronic-based speech training will prevail only on the assumption that the speech training system is well-designed. Not only that the technologies used must be adopted properly but it is also necessary that they serve the requirements of

the hard-of-hearing learners effectively. Here, we discuss some general requirements of such an e-training tool. The requirements, together with suitable approaches to the requirements, are presented in three aspects: the teaching, training, and tracking aspects.

**Teaching Aspect**

In order for learners to understand how to utter a sound correctly, they need to see and understand the production mechanism associated with that sound. In a traditional therapist-based situation, a student learns to produce a sound by imitating the action that he or she can observe visually. Most of the times, the observable action is limited to the shape of the lips. Physical limitation does not allow articulator inside of the mouth to be seen. Therefore, in teaching a learner to utter a sound, a computer-based instruction must be presented along with the visualization of related articulators. Three-dimensional computer graphics can be used for such visualization. The usage of three-dimensional models were adopted in [3] and [4], in which models of human head, the showing of whose facial skins can be toggled on/off, are used.

Dealing with children learners, the appearance of the three-dimensional human head model should be cartoon-like. Still, it must still resemble a real human head in the level that children learners can relate themselves with the models.

Prior to exposing children learners with the models, simple lessons on easy-to-identify articulators should be given so that children learners can understand the visualization.

**Training Aspect**

The system should provide computer-based exercises that let learners practice their speech production with minimal level of supervision. In such exercises, users are asked to say specified target utterances. Similar to every type of exercises, feedback should be given to the user. In this context, it means that the tools must be smart enough to analyze the speech uttered by the user in order to find out that the pronunciation of that speech is consistent with the pronunciation of the target utterance or not. In the traditional therapist-based situation, the exercises should be straightforwardly handled by the human specialist. However, it is far from simple in the case of computer-based one. Automatic speech recognition techniques can be used to evaluate the similarity between the uttered pronunciation and the target pronunciation. It can be deployed at different levels of pattern matching techniques.

Via a high-level perspective, we can group different level of pattern matching techniques for this task into two groups. Techniques belonging to the first group anticipate possible mistakes or pronunciation confusions in advance. For each test item, a group of imposter pronunciations are created and the user's speech, uttered in response to that test item, is compared with the correct pronunciation as well as the pronunciations of every imposters in order for the program to find the pronunciation that are best-matched with the user's speech. DP matching algorithm [7], Artificial Neural Networks [8], and Hidden Markov Models [9] are some examples belonging to this group. Techniques in the other group approach the problem differently. They do not anticipate possible mistakes in advance. Instead, user's speech is analyzed for key features determining phonemes containing in that speech. Techniques in this group would involve segmenting speech into a sequence of sound units, each of which is, in turn, analyzed for key features.

Techniques belonging to the second groups are more suitable to the speech training exercise than the ones in the first group since speech production knowledge

can be used heuristically and suggestions on how to correct the speech can be made based on such knowledge. However, not anticipating possible mistakes in advance makes the problem harder than doing so. An approach consistent with the second group was attempted in Leelaphatarakit et al. [10]

Another recommendation on the training aspect is that exercises should be presented in the form of games. This will motivate children learners to keep practicing.

**Tracking Aspect**

eLearning, as well as any types of learning, would be of no use if we cannot track the progress of the learners. The tracking module here may be thought of as an LMS (Learning Management System) in a typical eLearning environment. The main objective is to monitor students' progress from anywhere, at anytime. This capability is particularly important for our e-training due to the scarce resource of therapists for the hard of hearing. With this, the role of the teachers will be shifted to managing the results and to find ways to produce best possible outcomes, to design even more effective learning scheme, or even to tailor the courses to the needs of each hard-of-hearing student. A typical LMS tracks how much students have learned, how long they take to learn each lesson, and how well they understand the content. But for this work, the essence is to evaluate the obstacles that each hard-of-hearing student faces. The tracking module should be able to report the problem areas for each of the students such that attention can be put on the right place.

## 6. Conclusion

This paper describes our concept of an e-training system for assisting the hard of hearing children in their learning and practicing of how to utter Thai speech correctly. This ongoing work aims at developing both the framework and a prototype of such a system for the benefit of the Thai hard-of-hearing community in the near future.

## 7. References

[1] Soleymani, A.J.A.; McCutcheon, M.J.; and Southwood, M.H., **"**Design of speech illumina mentor (SIM) for teaching speech to the hearing impaired", in *Proc. of the 1997 Sixteenth Southern, Biomedical Engineering Conference*, 4-6 Apr 1997, pp.425-428.

[2] Massaro, D. M., "A computer-animated tutor for spoken and written language learning", in *Proc. of the 5th international conference on Multimodal interfaces*, Vancouver, British Columbia, Canada, 2003, pp.172-175.

[3] Kirschning, I.; Toledo, M. T.; Valadez, L.E.; Canizales, D., "Vowel & diphthong tutors for language therapy", in *Proc. of the Sixth Mexican International Conference on Computer Science*, 2005 (ENC 2005), Sept. 2005, pp.26 – 30.

[4] Bälter, O.; Engwall, O.; Öster, A; and Kjellström, H., "Designing for individuals with hearing impairment: Wizard-of-Oz test of ARTUR: a computer-based speech training system with articulation correction", in *Proc. of the 7th international ACM SIGACCESS conference on Computers and accessibility Assets '05*, MD, 2005.

[5] Jiang, X.; Wang, Y.; and Zhang, F., "Visual speech analysis and synthesis with application to Mandarin speech training", in *Proc. of the ACM symposium on Virtual reality software and technology*, London, United Kingdom, 1999, pp. 111-115.

[6] Hsiao, M.; Li, P.T.; Lin, P.Y.; Tang, S; Lee, T; and Young, S**, "**A computer based software for hearing impaired children's speech training and learning between teacher and parents in Taiwan",

in *Proc. of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2001, Volume 2, 25-28 Oct. 2001, pp.1457 – 1459.

[7] Osogami, Y. and Ishida, Y.**, "**A speech improver for hearing-impaired children using the DP matching and the SOM", in *Proc. of the 1996 IEEE TENCON, Digital Signal Processing Applications*, Vol. 1, 26-29 Nov. 1996, pp.:25 – 30.

[8] Zimmer, A.M.; Bingjun D.; Zahorian, S.A., "Personal computer software vowel training aid for the hearing impaired", in *Proc. of the 1998 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1998, ICASSP '98, Volume 6, 12-15 May 1998, pp. 3625 – 3628.

[9] Devarajan, M.; Fansheng M.; Hix, P.; and Zahorian, S.A., "HMM-neural network monophone models for computer-based articulation training for the hearing impaired", in *Proc. (ICASSP '03). 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003*, Vol. 2, 6-10 Apr. 2003, pp.II - 369-72.

[10] Leelaphatarakit, P., Punyabukkana, P., Suchato, A., "Locating Phone Boundaries from Acoustic Discontinuities Using a Two-staged Approach", in *Proc. of Interspeech2006*, Pittsburgh, PA, Sept.2006.