Three Dimensional Articulator Model for Speech Acquisition by Children with Hearing Loss

Arumugam Rathinavelu[1], Hemalatha Thiagarajan[2], and Anupriya Rajkumar[1]

[1] Dept of Computer Science and Engg, Dr.Mahalingam College of Engineering and Technology, Pollachi, TamilNadu, India
starvee@yahoo.com
[2] Dept of Computer Applications, National Institute of Technology, Trichy, India

Abstract. Our research indicates that acquisition of phonetic skills in voiced and voiceless speech sounds was improved by using Computer Aided Articulatory Tutor (CAAT). The interface of CAAT displays the place of articulation and relevant image objects for articulatory training simultaneously. The place of articulation was presented by using three dimensional articulatory tutor. Suitable computer graphics and Magnetic Resonance Imaging (MRI) techniques were used to develop inner articulatory movements of the animated tutor. Ten hearing impaired children between the ages 4 and 7 were selected and trained for 30 hours across four weeks on 50 words under 10 lessons. The words were selected from the categories of voiced and voiceless stops namely Bilabial, Dental, Alveolar, Retroflex and Velar. The articulatory performance of HI children was investigated to find out their speech intelligibility.

Keywords: 3D Modeling, MRI Techniques, Speech perception, Speech production, Computer aided articulator model.

1 Introduction

Many deaf children leave elementary school without required phonetic knowledge, which creates learning difficulties in their mainstream study. Many a time, children fail to follow the lip movement of teacher (1, 2). Inaccurate articulation is also considered as a speech defect (3). Most of the laboratory studies argued that auditory-visual speech perception is superior to visual or auditory perception alone. Three dimensional Synthetic head helps to enhance the intelligibility of audible speech (4). The previous experiments also indicated that the synthetic face can be used to transmit important visual speech information to hearing impaired children (5). The visualization of speech production helps the children to know about the place of inner articulators and to control his (or) her speech organs (6). Children hearing loss require constant guidance in articulation of such kind of speech sounds. Some of the distinctions in spoken language cannot be heard by children with degraded hearing. Due to hidden articulators and other social issues, visual speech perception is a complex task for hearing impaired children. Children with a moderate Hearing loss develop spontaneous speech, but their pronunciation often suffers from distortions and lack of articulatory precision. Our three dimensional vocal tract articulatory model has more visual

C. Stephanidis (Ed.): Universal Access in HCI, Part I, HCII 2007, LNCS 4554, pp. 786–794, 2007.

appearance and allows a realistic movement of the jaw, tongue and lips. Articulatory movements of Vocal Tract are simulated by employing suitable modeling and animation techniques. In India, most of the severely and profoundly hearing impaired rely on the visual modality alone for speech reception since speech and hearing training facilities are not readily accessible to them (7). The earlier mid-sagittal and 3D models focused on VT area research but none of them developed a complete articulatory model to train and improve the speech intelligibility of Hearing impaired children in Indian languages. Tamil is historically very old Indian language and articulating bilabial, dental, alveolar, Retroflex and Velar stops along with laterals and trill speech sounds are complex.

2 System Design

Hearing impaired children face the problem of distinguishing between the phonemes with place of articulation. The aim of our study was to develop a three-dimensional articulatory model based on a set of vocal tract geometrical data acquired by MRI on a reference subject [AR]. MRI is the dominating measurement method for three-dimensional imaging. No other method can compete with MRI today. Even though Electro-Magnetic articulography [EMA] and Electro-Palato Graphy [EPG] represent the kinematics in a limited way, it was found that MRI in the current status provides better overall kinematics of the oral cavity. So, MRI was selected as the basis for the 3D modeling and also the kinematics of articulations measured. MRI is non-hazardous for the subject also the reason of selection.

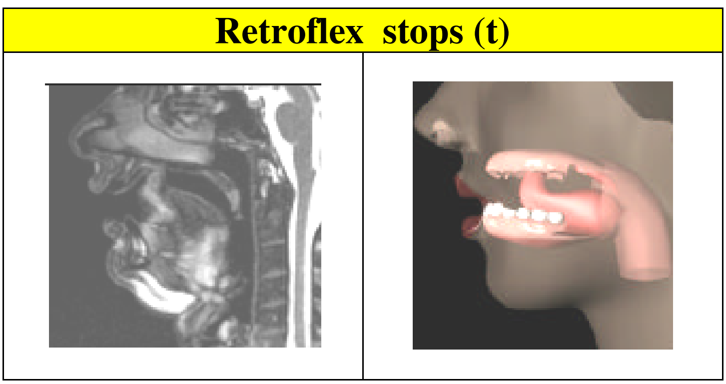

Fig. 1. (a) MRI (b) 3D Vocal Tract Model

The 3D articulatory model can be combined with a human face, creating a more familiar environment (8). Several researchers recognized the need to represent the tongue in a facial animation system (9). In our study, a realistic tongue was developed to achieve increased intelligibility among HI children (refer figure 1). Vocal tract information obtained from MRI scans (GE 1.5T Scanner) in the mid-sagittal, axial and oblique planes, which allowed us to construct 3D VT articulatory model. The 3D data are obtained from MRI of the subject, and front and profile video images of the subject's face. The tongue movements were built by animating tongue raise, tongue contact (with palate) and tongue curved.

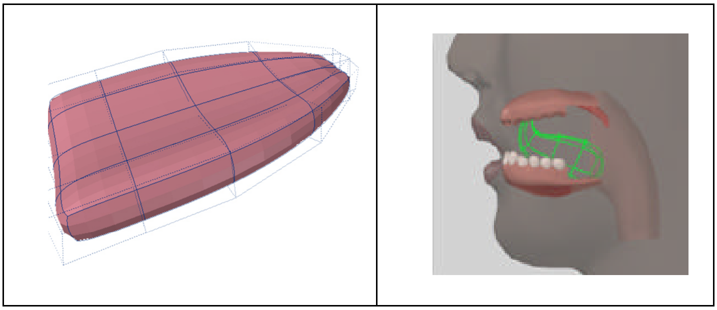

Fig. 2. (a) Parameterization of tongue (b) Tongue in Vocal Tract

The tongue model has been constructed from 50 control points. Five control points have been identified along the tongue from tip to back: two at the ends and three more to control the movement of the tongue tip, blade and dorsum. In the crosswise direction also five control points are located - two along each lateral and a medial point. The parameterization scheme chosen requires the medial position of the three lengthwise control points (excluding those at the ends) and two tongue lateral positions. The entire 3D shape of tongue was developed by using polygon mesh. Key frame techniques and interpolation between a finite set of visual targets were used to achieve speech articulation (11).

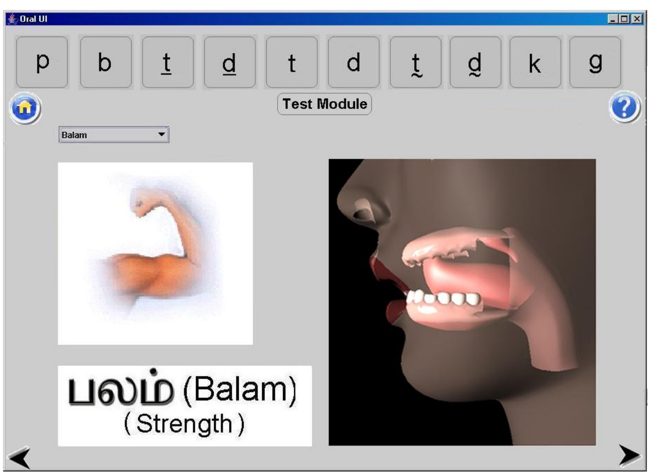

Fig. 3. The interface of the software model

The parameters used to construct 3D Vocal Tract model are a) jaw b) lips c) Tongue (tip, body, width) d) Velum (refer figure. 2). Natural speech of speaker was then synchronized as realistic audio visual sequence. The articulation of voiced speech sounds were indicated by coloring the vocal cord portion of model. In order to avoid the confusion in perceiving the similar speech sounds, visual cues were used in the software interface model to differentiate. For example, the visual cue 'palam' (fruit) was used for speech sound 'p' and the visual cue 'ball' was used for teaching syllables starting with 'b'. The upper-teeth were an obstacle to perceive the place of tongue position by the children and so were removed from the model. The inclusion of velum articulator is important to train the speech segments of Tamil stops. The graphical user interface of computer aided articulatory model was developed by using Java programming (refer fig.3). The necessary navigation controls were incorporated for moving between lessons, test module, training module, home, help, next and previous word. The user can select any one of the ten lessons by clicking on the button highlighting the sound and then choose a particular word from that category from the drop down list box. Once a word is chosen, the corresponding picture, the Tamil equivalent and its meaning are displayed on the left side of the interface. The right side of the panel displays the tongue position during the articulation of the sound, in the background of the human face. The tongue model has been developed using Java 3D. By employing the five parameters extracted from the MRI clipping and the value of the standard thickness of the tongue, the complete tongue model consisting of fifty control points is generated through suitable calculations.

3 Method

Five male and five female HI children aged between 4 and 7 were chosen for this study. All of them are profoundly deaf and they were given enough training to interact with Computer Aided Articulation Tutor (CAAT). There were three stages in this study: an initial test, training given by CAAT for 4 weeks and a final test. The combination of consonants [p, b, $\underline{t}$, $\underline{d}$, t, d, ***t***, ***d***, k, and g] and short and long vowels [a, e, I, o, and u] helped us to build 50 meaningful words under ten lessons for this computer aided articulation test (refer table.1). Each lesson consists of 5 words and five lessons allotted for voiced and remaining five for voiceless sounds. Every day, they spent 45 min each in fore-noon and after-noon sessions over the period of 4 weeks. Children were engaged by the computer aided articulatory tutor in acquisition of complex articulatory position of speech sounds placed in meaningful words at different contexts. In the interface of software model, based on the phoneme input and selection of corresponding word, the three dimensional tongue articulatory positions is shown along with Image object and title. After 4 weeks training, children were instructed to articulate the speech sounds and syllables in front of software model. The articulatory performance of children was investigated to find out in detail as follows a) Voiced lessons – Pre and Post-test b) Voiceless lessons – Pre and Post-test c) Lesson wise – Pre and Post test d) Gender wise.

Table 1. Place of articulation (Stops) [10]

Phoneme (Voiceless)	**Place of articulation (Stops)**	**Phoneme (Voiced)**
- p – puspam (flower), uppu (salt), appaa (father) - Lesson 2 -	**Bilabial-** In its production the lips are closed and the soft palate is raised to close the nasal passage. When the lips are opened the air suddenly comes out with explosion. Voiceless doesn't have vibration in the vocal cords.	**- B -** Balam (Strength), bimbam (Image), ambu (Arrow) - Lesson 1 -
- t̲ – tattu (plate), taamaray (lotus), Patthu (ten) - Lesson 4 -	**Dental-** It is produced when the tip of the tongue touches the upper teeth. The soft palate is also raised so that the air cannot get through the nasal cavity in its production. When the tip of the tongue is released from the upper teeth the air suddenly escapes through the mouth. There is no vibration in the vocal cords for voiceless sounds.	**- d̲ –** Darram (wife), pandam (torch), dikil (terror) - Lesson 3 -
- t – kattay (bundle), vetri (success), netti (forehead) - Lesson 6 -	**Alveolar-** In its production the tip of the tongue is placed against the alveolar ridge. The sides of the tongue are in contact with the teeth firmly. The soft palate is also raised so that the air cannot get through the nasal cavity. When the contact is released the air escapes through the mouth. There is no vibration in the vocal cords for voiceless.	**- d -** Andru (that day), indru (today), pandri (pig) - Lesson 5 -
- ṭ – tin (tin), ettu (eight), vattam (circle) - Lesson 8 -	**Retroflex-**It is produced by the tip of the tongue curved towards the back and making contact at the roof of the mouth. The soft palate is raised as in the production of other stops. There is no vibration in the vocal cords. When the contact is released the air escapes through the mouth.	**- d –** Kaday (shop), taaday (jaw), andam (world) - Lesson 7 -
- k – kadal (sea), kannadi (mirror), kay (hand) - Lesson 10 -	**Velar-** In its production the air stream is blocked by the back of the tongue while it is in firm contact with the soft palate. The soft palate is in raised position so that no air escapes through the nasal cavity. When the back of the tongue is released suddenly the air comes out of the mouth with explosion. There will be no vibration of the vocal cords during its production for voiceless.	**- g –** guru (teacher), Tangam (gold), ganapathi (Lord Ganesha) - Lesson 9 -

4 Results

Before computer aided articulatory training, the speech segments were presented to HI children to articulate those sounds first time. The mean and SD of Pre-test results were 40 and 6.57. Many of them have not performed well in Pre-test since the voiced and voiceless made them confused. For example, the phoneme 'B', '***D***', 'D̲', 'G', 'D','t' and 'k' were misarticulated due to its complex place of articulation as detailed in table 2. The misarticulated speech sounds were noted by us to focus more on those phonemes during training session.

Table 2. Pre-test results (Sample data)

Lesson No	Target Phoneme	Misarticulated
1	***B*** -alam (strength)	***P***-alam
	B -imbam (Image)	***P*** –imbam / **m** - ambam
2	***Pu*** -l (grass)	***Pa*** -l
3	Pan- **d̲** -am (torch)	Pat- ***t*** -am
	D̲ -aram (wife)	**T** -halam
4	Ta- ***t̲*** -tu (plate)	Te- ***r*** -ru
5	Pan- ***d*** **-**ri (pig)	Pat- ***t*** -ri
	In - ***d*** **-** ru (today)	In – **n** - ru
6	Net- ***t*** - ri (forehead)	Ne- r – ri, ne - n - ni
	Vet – ***t*** – ri (success)	Ver – r – ri, ve- n -ni
7	Ka- ***d*** –ay (Shop)	Ka – ***t*** - tai
	An – ***d*** –am (world)	At- ***t*** -am
8	Vat- ***t*** –am (circle)	Val – ***l*** -am
	Ti – n (tin)	***Ta*** -n
9	***G*** –uru (Teacher)	***K*** - ur/ uru/ ulu
	Tan – ***g*** –am (gold)	Tak- ***k*** -am
10	***K*** – adal (sea)	***A*** -adal
	K - ay (hand)	***Th*** - ai

The mean and SD of Pre-test results were 33 and 15.08 for voiced speech sounds and 36.5 and 10.98 for voiceless speech sounds (refer figure.4).

During four weeks training period, children were instructed to articulate the given list of speech sounds and meaningful words until they achieve maximum accuracy. Training time varied to individual based on their pre-test performance. As per our observation, most of the misarticulated speech sounds were corrected during training.

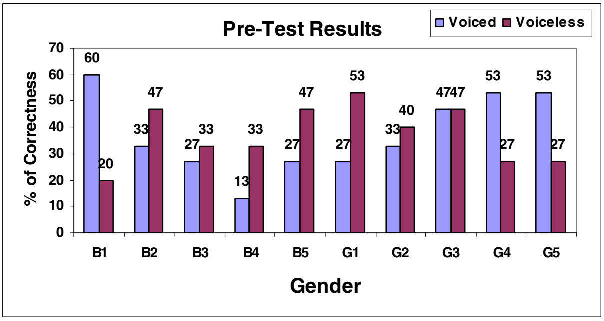

Fig. 4. Pre-test results (Voiced and Voiceless)

Table 3. Post-test results (Sample data)

Lesson No	Target Phonemes	Misarticulation
1	***B*** – imbam (strength)	***Ma*** - mbam
2	***P*** – akal (Day – time)	***Pa*** - lal
3	Pan- ***d*** -am (Torch)	Pa- ***tt*** -am
4	Tat- ***t*** –u (Plate)	Ta- ***d*** - u
5	Kan – ***r*** – u (calf)	Kan- ***n*** -u
6	Net – ***t*** –ri (Fore-head)	Ner – ***r*** -i
7	Ka- ***d*** –ai (Shop)	Ka- ***tt*** -ai
8	***T*** – in (Tin)	***T*** -en
9	***G*** – uru (Teacher)	**U** -ru
10	***K*** – adal (Sea)	K-a- **tt** - al

Most of their misarticulation was corrected during training and found satisfactory. Even after four weeks training, some of them have not performed well as per the details presented in table 3. The reasons for misarticulation may be due to voicing, individual capacity and acquisition time. The mean and SD of post-test results were 76.5 and 14.04. The mean and SD of Post-test results were 80 and 16.87 for voiced speech sounds and 73 and 17.20 for voiceless speech sounds (refer figure 5).

The overall performance of girls was better than boys in both pre- and post-test results (refer fig 6). In voiceless experiments (refer figure 6(b)), boys secured 36% (pre-test) and 61% (Post-test) against girls performance of 39% (pre-test) and 75% (Post-test). In voiced experiments (refer fig 6(a)), boys secured 32% (pre-test) and 75% (Post-test) against girls performance of 43% (pre-test) and 83% (Post-test).

The aim of computer aided articulation test was to increase the performance in articulation of speech sounds under every lesson. The pre-test and post-test results are shown in figure 7. The mean and SD of pre-test results are 37 and 4.81. The mean and SD of Post-test results are 75 and 9.38.

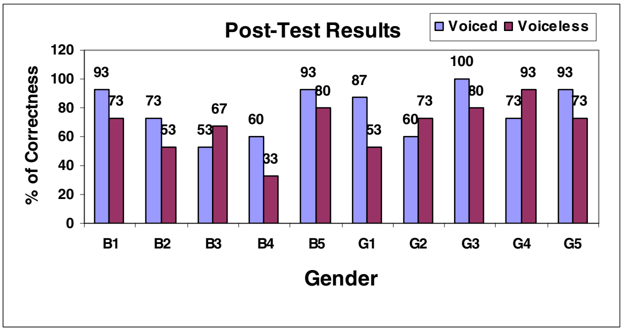

Fig. 5. Post-test results (voiced and voiceless)

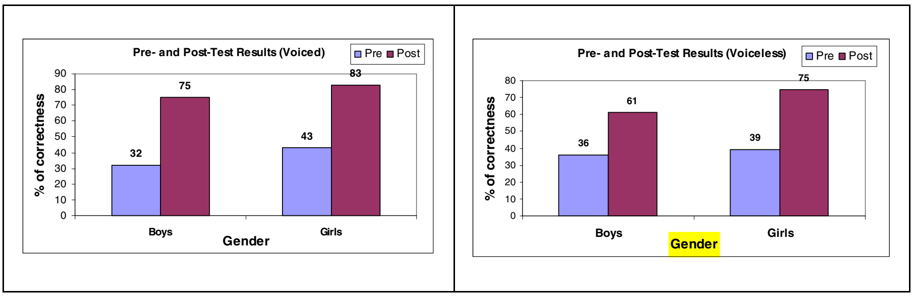

Fig. 6. Gender performance (a) in Voiced (Pre- and Post-test), (b) in Voiceless (Pre- and Post-test)

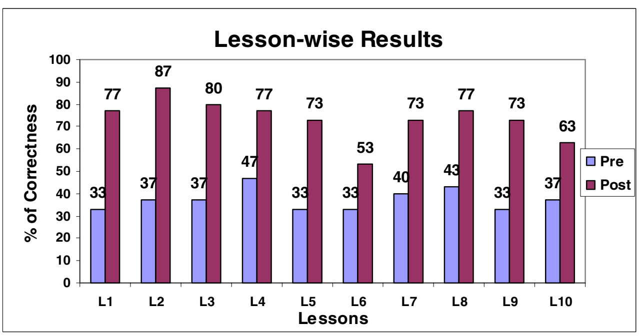

Fig. 7. Lesson-wise performance by children –Pre and Post-test results

5 Conclusion

The goal of our study was to examine the effectiveness of CAAT in teaching place of inner articulatory position of complex speech segments. Hearing impaired children were trained by computer aided articulatory model to perceive the place of articulation of speech segments. The misarticulated speech sounds were corrected during training period and the children improved the articulation to the accepted level in post-test performance. Realism of the visible speech is measured in terms of its intelligibility to the speech readers (12-16). This model succeeded in improving the speech intelligibility of hearing impaired children. This articulatory model is made available as an instructional tool for training syllables to HI children very effectively without involving any physical strains as human beings.

Acknowledgements. This research work is financially supported by Dr. Mahalingam College of Engg and Technology, Pollachi, South India. I wish to thank members of the college management and faculty colleagues of CSE and IT dept for their extended support.

References

1. Rathinavelu, A., et al.: E-Learning for hearing impaired. In: Proceed Apple Univ Consortium Conf 2001, Townsville, Austr. James Cook University pp. 21.1–21.6 (2001)
2. Rathinavelu, A.: Early language learning for elementary students with hearing impairment in India. Dissertation. Perth, Austr: School Communications Multimedia, Edith Cowan Univ. (2003)
3. Thirumali, M.S.: Language acquisition thought and disorder. In: CIIL Occasional Monographs Series - 10, Mysore, India (1977)
4. Fagel, S.: Merging methods of speech visualization. ZAS Papers in Linguistics 40, 19–32 (2005)
5. Siciliano, C., et al.: Lip readability of a synthetic talking face in normal hearing and hearing impaired listeners. In: Dept of Phonetic and Linguistics, University College London (2003)
6. Vicsi, et al.: A Multilingual teaching and training system for children with speech disorders. Intl. J of Speech Tech 3, 289–300 (2000)
7. Thirumalai, M.S, Gayathri, S.G: Speech of the Hearing Impaired. In: CIIL Occasional Monographs Series -43, Mysore, India (1998)
8. Engwall, O.: Modelling of the vocal tract in three dimensions. In: Centre for Speech Technology (CTT), KTH, Stockholm, Sweden (1999)
9. King, S., et al.: Creating Speech-Synchronised Animation. IEEE Transactions on visualization and computer graphics 11(3), 341–352 (2005)
10. Rajaram, S.: Tamil Phonetic Reader. In: CIIL, Mysore, India (2000)
11. Bailly, G., et al.: Audio Visual Speech Synthesis. In: Int J of Speech Tech, Netherlands, pp. 331–346. Kluwer Academic Publishers, Boston (2003)
12. Rathinavelu, A., et al.: Evaluation of a computer aided 3D lip sync instructional model using VR objects. Int J Disabil Hum Dev. 5(2), 127–132 (2006)
13. Rathinavelu, A., et al.: Interactive multimedia tool to help vocabulary learning of hearing impaired children by using 3D VR objects as visual cues. Nat J Technol. 2(1), 25–32 (2006)
14. Rathinavelu, A., et al.: Computer aided articulation tutor using three dimensional visual cues for the children with hearing loss. J Comput Sci. 2(1), 76–82 (2006)
15. Rathinavelu, A. et al.: Evaluation of a computer aided 3D lip sync instructional using virtual reality objects. In: Proc. 6th Intl Conf. Disability, Virtual Reality & Assoc. Tech, Esbjerg, Denmark, pp. 67–74 (2006)
16. Massaro, D.W.: Perceiving talking faces: From speech perception to a behavioral principle. MIT Press, Cambridge (1998)