computing@tanet.edu.te.ua
www.tanet.edu.te.ua/computing

ISSN 1727-6209
International Scientific
Journal of Computing

Computer Aided Articulatory Tutor: A scientific study

Arumugam Rathinavelu [1], Hemalatha Thiagarajan [2]

[1] Dr. Mahalingam College of Engg & Technology, Pollachi, South India,
starvee@yahoo.com, www.drmcet.ac.in
[2] National Institute of Technology, Trichy, South India

Abstract: *This paper describes the use of Computer Aided Articulation Tutor (CAAT) to conduct phonetic training to the hearing-impaired (HI) children with inner articulators as visual cues. This Articulatory tutor was developed by using Magnetic Resonance Imaging (MRI) and computer graphics techniques. The articulators include the movement of jaw, lips, tongue and velum. Ten hearing impaired (HI) children between the ages 4 and 7 were selected and trained for 10 hours across 4 weeks on 24 words. Intelligibility of HI children was investigated to find out their performance in speech perception and articulation. The post-training results indicated that HI children improved in articulation of speech sounds placed at different contexts. This 3D visual simulation tool helped HI children in perceiving the speech information significantly.*

Keywords: *3D Modeling, MR Images, Speech perception, Speech production, Computer aided articulatory model*

1. INTRODUCTION

An articulation disorder of hearing impaired children may be defined as incorrect production of speech sounds due to faulty placement, timing, direction, pressure, speed or integration of the movements of lips, tongue, velum or pharynx. Once a child has a reasonable command of language and his phonetic and phonologic skills enable him to produce most speech patterns [1]. According to deaf researchers [2], hearing children have consistently demonstrated the ability to perform such tasks between the ages of 4 and 5 years. But hearing impaired children are delayed by approx. 3 years in this cognitive developed milestone. According to empirical findings, children are good at producing spoken language if they do better at speech perception. When language and articulation disorders coexist, treating only one of them may produce some effect on the other, but it is likely that the effects will not be substantial; more research is needed [3]. The animated face model of SAKTHI visualizes the articulator movements as described by the articulation parameters like lips, jaw and tongue [4]. But no study was carried out by using computer aided articulatory model to teach articulator position of speech segment of Dravidian language syllables. Tamil is historically very old Dravidian language and articulating three types of laterals and trill speech sounds are little complex. Many deaf children leave elementary school without required phonetic knowledge, which creates learning difficulties in their mainstream study. Deaf children perceive language from teachers' lip-movement during their classroom study hours. Many a time, children fail to follow the lip-movement of teacher [5, 6]. According to researchers [7], complementarily of auditory and visual information simply means that one of the sources is most informative in those cases in which the other is weakest. For example, the difference between /ba/ and /da/ is easy to see but relatively difficult to hear. On the other hand, the difference between /ba/ and /pa/ is relatively easy to hear but very different to discriminate visually. Some of the distinctions in spoken language can't be heard with degraded hearing. Inaccurate articulation which is wholly or partly unintelligible is also considered as a speech defect [8]. Most of the laboratory studies are of the opinion that auditory-visual speech perception is superior to visual or auditory perception alone. 3D Synthetic head helps to enhance the intelligibility of audible speech [9]. The previous experiments also indicated that the synthetic face can be used to transmit important visual speech information to HI children [4, 10]. The visualization of speech production helps the children to know about the place of inner articulators and to control his (or) her speech organs [11]. 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 [12]. The amount of speech perception increases by using 3D visual cues [4, 13]. The poor articulation of HI children was improved by using inner articulators of an animated tutor in our earlier studies [14].

2. SYSTEM DESIGN

Hearing impaired children face the problem of distinguishing between the phonemes with place of articulation. Articulatory position of Tamil laterals and trills is complex even for normal hearing children. Due to the child's problems hearing the mispronunciation, the therapist uses multimodal techniques (vision, touch, and, if possible, hearing) to show where and how the articulation is produced and how different speech sounds differ from each other [15].

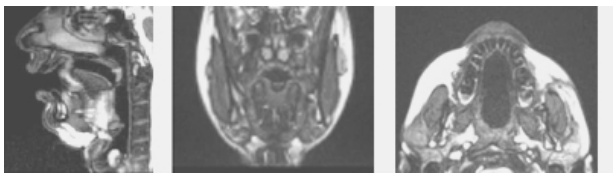


Fig. 1 – MR Images for the phoneme 'la' in three planes

The visual modality can complement the auditory information providing segmental cues on place of articulation and prosodic information concerning prominence and phrasing [16]. The 3D model can be combined with a human face, creating a more familiar environment [17]. The tongue plays an extremely important role in complex place of articulation and it helps to produce the desired speech. Several researchers recognized the need to represent the tongue in a facial animation system [18]. Due to hidden movement of tongue, the talking face needs transparency for visual representation.

2.1 MRI TECHNIQUES

Internal geometry of human (VT) can not be retrieved from Video-image processing. Computed Tomography (CT) and MRI are the two methods for providing full 3D data of the entire VT. Compare to CT, MRI is considered as non-invasive for the subject. Vocal tract information obtained from MRI scans (GE 1.5T Scanner) in the sagittal, axial and coronal planes, which allowed us to construct 3D VT articulatory model (fig. 1). The data of 3D Vocal Tract model was retrieved from MRI and Video footage of native male speaker of Tamil [AR]. The 3D inner articulatory model was created with correspond to the visemes that constitute a word. The tongue movements were built by animating tongue raise, tongue contact (with palate) and tongue curved.

2.2 VT SIMULATION MODEL

An articulated model is a collection of many deformable objects like jaw, lips and tongue connected together. The parameters used to construct 3D VT model are a) jaw b) lips c) Tongue (tip, body, width). Video clips of Tamil native speaker were shot and the frames of this footage provided a guide for the corresponding frames of an animation [19]. A 2D surface based coordinate grid was mapped onto the front and side images of a face. Point correspondences were established between the two images and the grid was reconstructed in 3D space [20]. The 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 [21]. In order to get the parameterized tongue model, a common geometric representation is defined and then animated for each given phoneme. The tongue is modeled as a set of polygons. The tongue is visualized as being made up of two layers each containing 25 control points arranged in a 5 x 5 grid (fig. 2). In order to parameterize the tongue, five major control points have been identified. These include the tongue center (c), two control points along the right (viewed from the tongue tip) lateral median (l1 and l2) and the lower two control points along the vertical center line (p1 and p2). All the five control points are located on the top layer of the tongue.

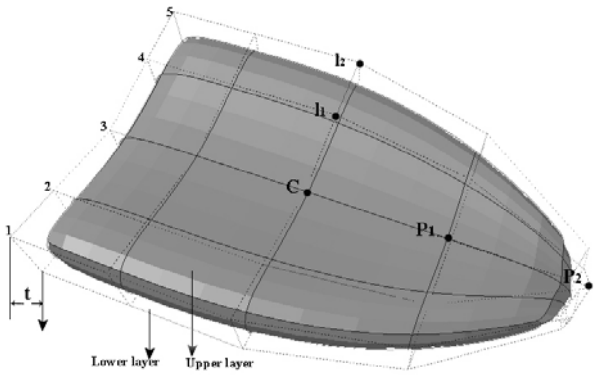
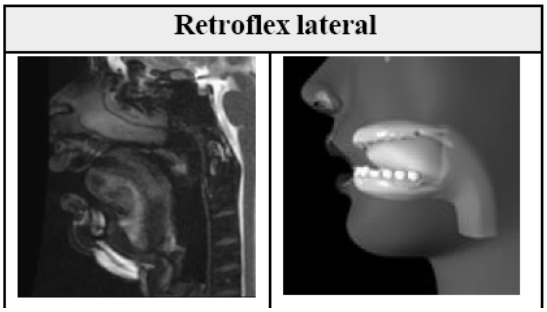


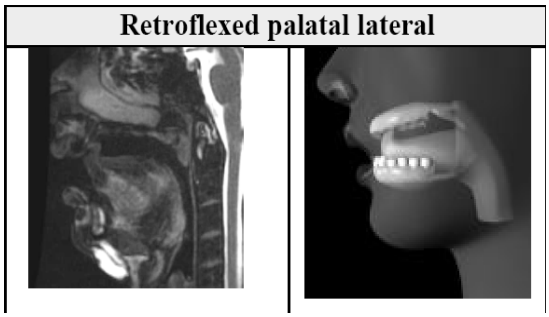
Fig. 2 – Tongue Model

The articulation of different phonemes results in different tongue positions. In order to model this, the five control points for the tongue are extracted from the mid-sagital MRI images captured during the articulation of each phoneme. These values are stored in a database along with the corresponding phoneme. Other data required include the standard thickness of the tongue – t and a correction factor x. The correction factor is used because the tongue tapers towards the tip. The interface of Computer Aided Articulatory Tutor (CAAT) displays both the

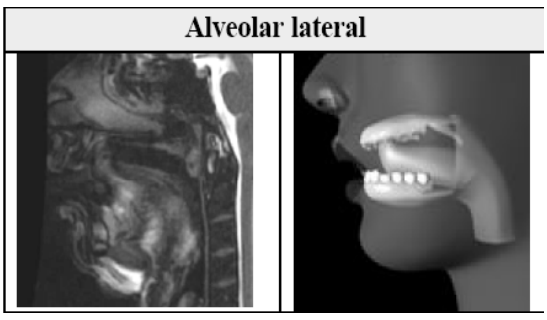
articulatory position of each word and 3D and 2D view of corresponding object together. The deformation of tongue model presents the place of articulation under each category of speech sound along with pictures. The graphical user interface of computer aided articulatory model was developed by using Java programming. The necessary navigation controls were incorporated for moving between lessons, test module, training module, home, help, next and previous word.



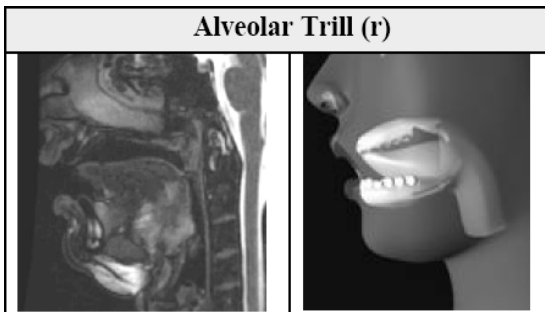
(a) (b)
Fig. 3 – Retroflex lateral (ɻ)
(a) MRI and (b) 3D VT



(a) (b)
Fig. 4 – Retroflex palatal lateral (ɭ́)
MRI and (b) 3D VT



(a) (b)
Fig. 5 – Alveolar lateral (l)
(a) MRI and (b) 3D VT



(a) (b)
Fig. 6 – Alveolar Trill (r)
(a) MRI and (b) 3D VT

3. EXPERIMENTS

In this experiment, the inner articulatory movement of vocal tract (tongue, velum and palate) was used to train children with hearing impairment (HI) on articulation. The purpose of this experiment was to see whether HI Children improve articulation by looking at the mid-sagittal view of lower face (figure 3-6). This view includes the movements of lips, jaw, tongue and velum.

Subjects: Ten HI Children in the age range of 4 - 7 years participated in the study. All of them had profound hearing loss and there were 5 boys and 5 girls. All children had misarticulation of laterals and trill.

Material: Twenty-four Tamil words with three laterals and one trill in various contexts were used. There were 4 sets of words with 6 words in each set (Table 1). The 3D Vocal Tract (VT) Simulation view of the lower face was used to train HI children on articulation of laterals and trill. As described in [22], laterals are pronounced by a complete blockade of air in the middle of the mouth but leaving a free passage for the air to pass through both sides of the tongue. Trill is produced by the rapid vibrations by the tip of the tongue against the alveolar ridge.

Procedure: Children underwent a pre-training test, training and post-training test. In pre-training test, articulation of trill and laterals was assessed by teachers. Training was for 4 weeks with 30 minute session every day. On the whole they were trained by 3D VT Simulator for a total of 10 hours in 20 days. Following training, children's articulation of all these words were recorded and tested by their teacher for articulation. Each correct articulation was given a score of '1'.

Table 1. Description of speech sounds of category and sample words

Sr. No	Category	Description	Sample words
1	Alveolar Lateral - l	The lip of the tongue has contact with the alveolar ridge in such a way that there is complete blockade of air in the middle of the mouth. The air is allowed to pass by the sides of the tongue since they are not in contact with the sides of the palate. The vocal cords are vibrated during its production [22].	cheval (Cock) balam – (Strength) ka:lam – (Time) ka:l – (Leg) eli – (Rat) lattU – (Lattu)
2	Retroflex Lateral - l̲	The tip of the tongue is slightly curved and made to contact the middle of the palate. The air stream is completely blocked in the middle of the mouth. The vocal cords are vibrated during its production [22].	uli – (Chisel) thavalai– (Frog) maylam– (Drums) manjal- (Turmeric) pallam – (Pit) palli – (School)
3	Palatal Lateral - Í	The tongue is curled back and the tip of tongue is placed very near the roof of the mouth but not touching it. The air stream is allowed to pass through the sides of the tongue as well as in between the tip of the tongue and the roof of mouth. The vocal cords are vibrated [22].	malai – (Rain) kuli – (Pothole) ko:li – (Hen) Va: lai – (Banana) kalugu – (Eagle) Palam – (fruit)
4	Alveolar Trill - r	It is produced by the rapid vibrations by the tip of the tongue against the middle of the alveolar ridge. The soft palate is raised to close the nasal passage. The vocal cords are vibrated [22].	maram – (Tree) nari – (Jackal) kurangu- (Monkey) karuppu –(Black) erumbu– (Ant) rambam- (backsaw)

4. RESULTS AND DISCUSSION

Three types of lateral and one trill were used in our experimental study. Figure 7 shows the performance of individual subjects on retroflex lateral. Pre-test mean score of 2 (SD = 0.48) and Post-test mean score of 5 (SD = 0.67) were obtained for retroflex lateral. All of them improved articulation in post-training test.

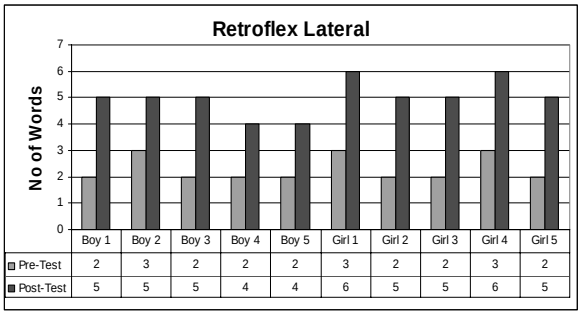


Fig. 7 – Pre- and Post- Training test Scores of children on retroflex lateral.

Figure 8 shows the performance of individual subjects on retroflexed palatal lateral. Pre-test mean score of 2 (SD = 0.48) and post-test mean score of 4.5 (SD = 0.53) was obtained for retroflexed palatal lateral. An improvement in articulation was observed in post-training test.

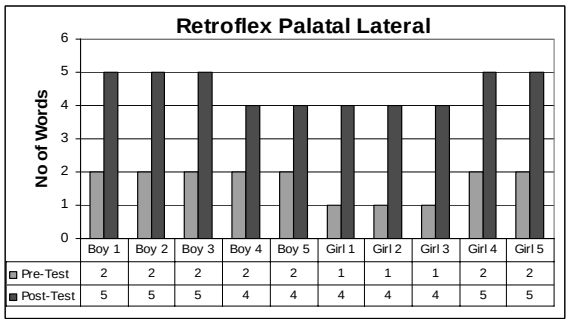


Fig. 8 – Pre- and Post-training test scores Of children on retroflex palatal lateral

Figure 9 shows the performance of individual subjects on alveolar lateral. Pre-test mean score of 4 (SD= 0.48) and post-test mean score of 5.5 (SD= 0.53) were obtained for alveolar lateral.

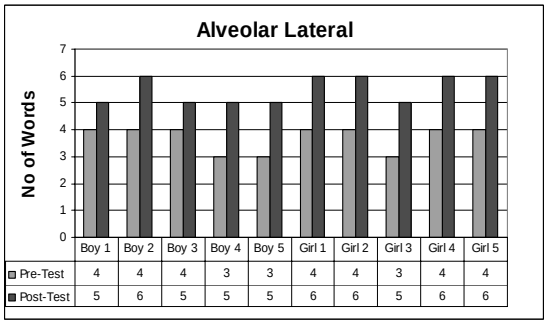


Fig. 9 – Pre- and post-training test Scores of children on Alveolar lateral

Pre-test mean score of 1 (SD = 0.42) and post-test mean score of 3 (SD = 0.42) were obtained for alveolar trill. Figure 10 shows the performance of individual subjects.

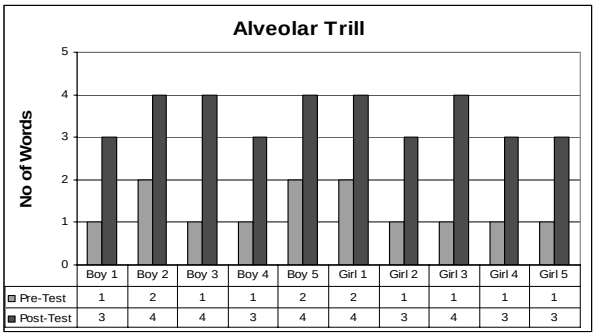


Fig. 10 – Pre- and post-training test Scores of children on alveolar Trill

Post-test results (Figure 11) showed that over all, the children improved their articulation (Mean: 18.5, SD: 1.35) by the training of 3D VT Simulator, in perceiving the place and manner of articulation of phonemes of each word. Figure 11 shows the pre- and post-training scores on all four speech sounds.

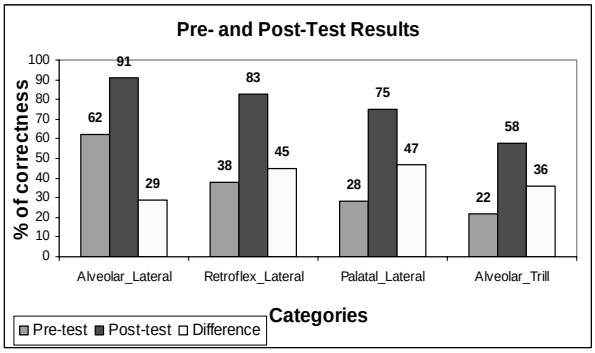


Fig. 11 – Pre- and Post- test Results

The misarticulated words (Pre-test mean score of 9 and SD = 1.15) were corrected during training period and the children improved to the accepted level of speech production in post-test performance.

5. CONCLUSION

A variety of training programs have been advised to Indian deaf children as instruction aid in acquisition of speech segments, but no study reported as such using three dimensional computer aided articulatory model as aid to deaf in the acquisition of spoken language. The goal of our investigation was to design, develop and evaluate a computer aided articulated tutor for children with hearing loss. In our experimental study, we examined whether the children with hearing loss could able to perceive the articulatory position from 3D VT model and then articulate the same during training and in final test. This 3D Vocal Tract model provided more visual information on articulatory position of complex speech sounds. The results of our study indicated that subjects' performance was best on alveolar lateral and lowest on alveolar trill. The results also indicates that 100% score was not obtained on any phoneme especially post-training test score was low on alveolar trill | r |.

6. REFERENCES

[1] Ling, D, Speech and the hearing impaired child: Theory and Practice, Alexander Graham Bell Association for the deaf, Inc, U.S.A, 1976.

[2] Lundy, J.E.B., Age of language skills of deaf children in relation to theory mind development, Journal of deaf studies and deaf education 7:1, 2002, pp.41-56.

[3] Pena-Brooks, A and Hegde, M.N, Assessment and treatment of articulation and phonological disorders in children, Pro-Ed Publishers, 2000.

[4] Rathinavelu, A et al., Computer Aided Articulation Tutor using Three dimensional Visual cues for the children with hearing loss, Journal of Computer Science (ISSN - 0973-292 6), Vol.2, No.1, 2006, pp 76-82.

[5] Rathinavelu, A and Gowrishankar, A., e-Learning for hearing impaired, Proceedings of the Apple University Consortium Conference 2001 (ISBN 0-947209-33-6), James Cook University Townsville, Queensland Australia, Sept 23-26, 2001, pp 21.1-21.6.

[6] Rathinavelu, A., Early language learning for elementary students with hearing impairment in India, Masters' Thesis, School of Communications and Multimedia, Edith Cowan University, Australia, 2003

[7] Massaro, D.W., & Light, J., Using visible speech for training perception and production of speech for hard of hearing individuals. Journal of speech, Language, and hearing research, 47(2), 2004, pp.304-320.

[8] Thirumali, M.S., Language acquisition thought and disorder, CIIL Occasional Monographs Series - 10, Mysore, India, 1977.

[9] Fagel, S., Merging methods of speech visualization, ZAS Papers in Linguistics 40, 2005, pp.19-32

[10] Siciliano, C et al., Lip readability of a synthetic talking face in normal hearing and hearing impaired listeners, Dept of Phonetic and Linguistics, University College London, 2003.

[11] Vicsi et al., A Multilingual teaching and training system for children with speech disorders, Intl. Journal of Speech Technology, 2000, Vol.3, 289-300.

[12] Thirumalai, M.S and Gayathri, S.G., Speech of the Hearing Impaired, CIIL Occasional Monographs Series - 43, Mysore, India, 1998.

[13] Rathinavelu, A et al., Interactive multimedia tool to help vocabulary learning of hearing

impaired children by using 3D VR objects as visual cues, National Journal of Technology (ISSN-0973-1334), Vol.2, No.1, 2006, pp.25-32.

[14] Rathinavelu, A., Hemalatha, T and Savithri., Evaluation of a computer aided 3D lip sync instructional model Using VR Objects, ICDVRAT, Sept 18-20, Denmark, 2006, pp 67-73.

[15] Eriksson et al., Design Recommendations for a computer-based speech training system based on End-user Interviews, Centre for Speech Technology (CTT), KTH, Stockholm, Sweden, 2004.

[16] Granstrom, Multimodal Speech Synthesis/NGSLT-Speech Technology Course, Centre for speech Technology (CTT), KTH, Stockholm, Sweden, 1999.

[17] Engwall, O., Modeling of the vocal tract in three dimensions, Centre for speech Technology (CTT), KTH, Stockholm, Sweden, 1999.

[18] King, S et al., Creating Speech-Synchronized Animation, IEEE Transactions on visualization and computer graphics, Vol-11, No.3, May/June 2005, pp. 341-352.

[19] Lewis, J. Automated Lip-Sync: Background and Techniques, The Journal of Visualization and Computer Animation, 2:118-122, 1991.

[20] Parent, R, Computer Animation Algorithms and Techniques, Morgan Kaufmann Publishers, San Francisco, 2002

[21] Bailly, G et al , Audio Visual Speech Synthesis, International Journal of Speech Technology, Kluwer Academic Publishers, Netherlands, 2003, pp.331-346.

[22] Rajaram, S., Tamil Phonetic Reader, CIIL, Mysore, India, 2000.

Arumugam Rathinavelu *received his Masters degree from school of communications and Multimedia, Edith Cowan University, Perth, Australia in the year 2003. Currently, he is working as Asst. Professor and head of the dept of Computer Science and Engg at Dr.Mahalingam College of Engg and Technology, Pollachi, Tamilnadu, South India. He has over 15 years of academic and industrial experience. He is currently pursing his PhD at National Institute of Technology, Trichy, South India. He is a member of ACM, Computer Society of India and Indian Society of Technical Education. His area of research interest includes computer graphics and animation, multimedia technologies, Web engg and Speech sciences. He has published over 15 papers at various national and international conferences and journals.*

Dr. Hemalatha Thiagarajan *received her PhD from University of Texas, USA. She has over 30 years of academic and research experience. She is now working as Asst. Professor in Computer Applications and as Associate dean – student services at National Institute of Technology, Trichy, South India. Her area of specialization includes Operations research, Algorithms, Neural Networks and Image processing. She has published over 15 papers at various National and International Conferences & Journals.*