An experiment using personalised multimedia interfaces for speech therapy

Jennifer George[1] and Paul Gnanayutham[2]

[1] SAE Institute, United House, North Road, London, N7 9DP,
United Kingdom, +44 (0) 20 7609 2653, jennifer.george@sae.edu
[2]Department of Computing, University of Portsmouth, Buckingham Building
Lion Terrace, Portsmouth PO1 3HE, United Kingdom
+44 (0) 23 9284 6404, paul.gnanayutham@port.ac.uk

Abstract. This paper discusses the experiments used in the design and development of assistive technology interfaces for children with phonological disorders and answering the research question *should there be a group of interaction paradigms or one novel interaction paradigm that can be personalised, to enhance the performance of pronunciation skills for children with speech impairments?* This research question is tested against an artifact designed and developed to address the need for an assistive device to rehabilitate children with phonological disorders.

Keywords. Speech Impairments, Phonology, Special Needs, Interfaces, Speech Therapy, Methods, Methodology, Articulation and Assistive Technology.

1   Introduction

Children are often forced into mainstream schools, while special needs schools are shut down [1]. Children with phonological disorders go through fear, shame and frustration of not being understood and being misunderstood. The special educational needs committee recommended that this group of children must be given the highest priority which should be the hallmark of a good education system and a civilised society. In cases where parental or professional support is lacking, arguably it is useful and effective to have interactive computer devices that could teach, train and guide the child at his or her own pace. This study investigated the possibilities of using a three dimensional interface to explain the anatomy of articulating and to facilitate children with speech impairments to articulate better. Three-dimensional applications have the ability to explain visually the organs of speech used in formation of clear sounds. The research addressed this issue by way of designing and developing a multimedia based assistive device that helped special needs children in mainstream education improve pronunciation.

2   Background

An extensive literature search into methodologies and methods used in the areas of speech Impairments, Interfaces, Speech Therapy, Pronunciation, Articulation, Human Computer Interaction and Assistive Technology was carried out. It took into consideration psychological issues, educational methods, ethical issues, developmental practices and design methodologies that were utilised to create and evaluate assistive devices for this purpose.

When the air released from the lungs passes through the mouth, according to the organs of speech used in formation the different sounds and words are formed. There may be both mental and physical difficulties in achieving this. Disordered speech may occur due to poor auditory discrimination, difficulty in forming sounds, structural difficulties, verbal dyspraxia or hearing difficulties [2]. The child's chronological age and developmental level may differ according to the mental and biological growth of each child. The target audience of the research included disabled children and as these children had varied mental ages, no age limit was applied to chronological age. Verbal information does not primarily have to be caused by auditory or speech production difficulties. This can also be associated with a relatively general memory deficit or be due to poor auditory skills [3].

There are various conventional strategies, methods and games used to work on speech, sound and phonological problems. Initially, the speech and language therapist identifies the sounds that need therapy. The therapy can involve encouragement of articulators, by way of making play sounds (eg. Beep beep, brrrm), making animal sounds and sounds for action songs. The Jolly Learning is a system widely used by speech therapists in the UK. Jolly Phonics is an area in the Jolly Learning system that works with phonological disorders [4]. Actions from this are used to demonstrate and practice sounds in traditional types of therapy. Other therapeutics activities may also include auditory input therapy, structured listening or focused auditory input [5]. The animations included in the proposed application follows the Jolly Phonics animation along with their classification.

Various AAC accessories based on multimedia interfaces have been developed and are being used [8]. Children with severe communication impairments use alternative communication modes such as manual signs and visual symbols, which is a substitution for spoken words [9]. However, most of these devices include a speech recognition element. When children have difficulty in pronouncing clearly it is not possible for these devices to assist them.

Gong [10] stated that human appeal, liveliness, intuitiveness, anthropomorphic representations such as talking heads, pictorial characters and voices have become prevalent in computer interfaces. Gong argues that mixing a human face with a synthetic voice or synthetic face with a human voice creates inconsistency and violates users schema of what is human and what is humanoid. The proposed artifact uses humanoid face with human voice to demonstrate pronunciation of different sounds, as it is not possible to produce a video of the movement of the organs of speech from inside the oral cavity. Three-dimension (3D) has been chosen as the media to display the organs used in formation for every sound as realistic as possible accompanied with both male and female human voices.

Meyer [11] described various issues when writing software for children with speech impairment. One of the suggestions made by Meyer is using photos or realistic images that will be beneficial to the child. Careful attention should also be paid to restrict the number of images between four and eight. Reinforcement of sounds and words is also advised and any musical interlude and movement will also keep the child interested during learning. The proposed interface contains two animations and an image. The 3D property represents parts of the mouth and their functions in producing sounds as realistic as possible without scaring the child. Reinforcement is an optional facility provided with a replay button.

Watanabe and colleagues [12] developed a speech driven embodied interactive actor called InterActor with both functions of speaker and listener in activating human interaction and communication by generating expressive actions and motions related to speech input. InterActor is a robot, which listens and talks. This was a 3D interface that enhanced face-to-face communication and non-verbal actions and motions. The 3D property in the proposed interface acts as a speaker nut it is beyond the scope of this project to have a listening or speech input device.

Bälter and his team [13] also conducted a study for children with language impediments using the Wizard-of-Oz for testing a computer–based speech training system with articulation correction called ARTUR. ARTUR used audio-visual detection of mispronounced speech, marker-less tracking of facial features from video, articulatory inversion, articulatory model, and adaptation of the model to the user and feedback display. The Wizard-of-Oz study was used for developing the speech technology components and the interface of this system. The proposed interface has a replay button that facilitated this through a replay button. Children aged 9-14 gave positive remarks and described ARTUR as very intelligent and good. Children aged six needed more prompting and needed help from speech therapists. The therapists and carers can decide the important sounds or the group of sounds that need training for the younger children and the older children with better cognitive skills may be able to decide for themselves.

Lukaszewicz [14] presented a new concept of computer user interface based on recognition of ultrasound image of a tongue to develop a speech output. Various ultrasound images of tongue where stored in a computer with the sound associated with each image. When a disabled user attempted to speak ultrasound images were recorded and tongue images compared with the stored images to output the pre-recorded sound. The proposed interface attempts to display the production of sounds using almost realistic organs of speech. This decision was made, as it is less complicated and less scary for the children but at the same time fulfils the purpose.

3   Methodology

The design, development and experiments were carried out in two cycles and two phases (Fig. 1). The cycles were iterative processes that made the final versions of the group and personalised interfaces possible prior to conducting the experiments [6]. Phase one experiment was conducted on the group interface to explore *"Should there be a group of interaction paradigms to enhance the performance of*

*pronunciation skills for children with speech impairments or…"*and phase two experiments were conducted with the personalized interfaces to explore *"….should there be one novel interaction paradigm that can be personalised to enhance the performance of pronunciation skills for children with speech impairments."*
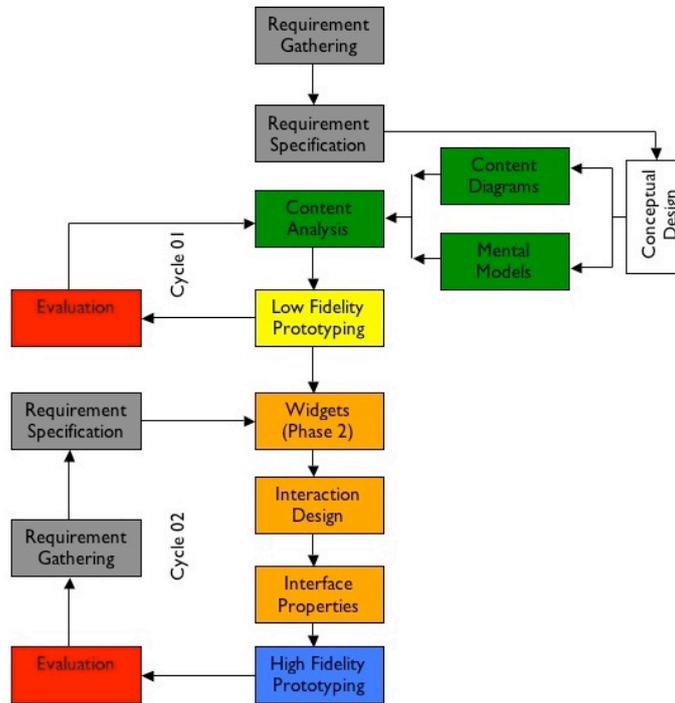
Fig. 1. Methodology

4   Design and Development

A focus group of ten participants was used for evaluation during design and development. The prototype was built using Flash. The interface developed in Phase 1 had an intro page, main page (Fig. 2) and an exit page whereas; the second interface had two additional pages with preferences (Fig. 3) and help. The intro page had a simple animation that would welcome the user. The main page would have the sounds available together with the front view and three-dimensional model animation [7] and another animation [4] that would help the user relate to the sound. The exit page would simply say goodbye. The preferences page gave an option of choosing if the user wanted all the sounds, background colour preferences, what animations they wanted to view and gender of sound [5]. The help page explained the terms used in the Preferences page.
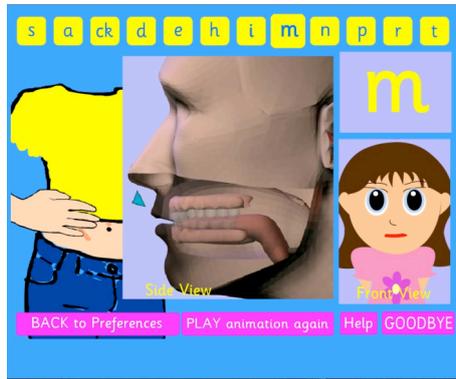
Fig. 2. Main Interface

Fig. 3. Preferences Interface

5   Experiments

The outcomes of both phases were to measure the success rate of the interfaces. The results were obtained by taking into account:
1. The number of attempts needed to pronounce each sound
2. The amount of training needed
3. The sounds that cause difficulty

Apparatus: Computer with CD-Rom and speakers, interface program

Experiment 1 was the control experiment where the participant was asked to attempt pronouncing the chosen sounds without the help of any interface. Experiment 2 was conducted with a general interface and Experiment 3 with the support of a personalised one. In Experiments 2 and 3, the functionality of the interface was explained to the participant by way of demonstration. It had to be explained to the participant that the 3D model in the interface was what the inside of the mouth looked like. Then the participant was asked to imitate the 3D animation for each sound.

Maximum of ten attempts were given for each sound. No pre-requisites or training was necessary to carry out the experiments. Participants between ages 5 years to 36 years, 4 females and 6 males, 7 able and 3 disabled evaluated the interfaces.

5.1   Experiment 1

No training was required to carry out this experiment. Difficulties in pronouncing sounds were taken note of. Participant 6 had difficulty in repetitive attempts. Participant 7 was enthusiastic but had difficulty understanding explanations. Overall difficulty was found in pronouncing s, ck, h, n and r. Participant 10 had specific high frequency sounds such as a and m but found others confusing and difficult. Difficulty was also noted in distinguishing t from d and n and r. The next experiment will be conducted using the developed interface to investigate whether further progress can be made in pronunciation for this group of participants.

5.2   Experiment 2

This phase of the study investigated whether disabled participants can be grouped together, when developing interfaces and users can use Group Interface with minimal training.

No training was required to carry out this experiment. Users couldn't be grouped as no specific patterns could be identified. T-tests are conducted to analyse the significance between two measured samples. It uses the standard deviation of the samples to estimate the standard error of the sampling distribution. This test is particularly useful when only a small sample is available. T-test was conducted on the two sets of data (without any interface and with group interface). The t-test shows that the two sets of data are significantly different at the 0.1% level (0.000013). Hence it can be deduced that using the group interface is 99.9% better than using no interface.

5.3   Experiment 3

The participant was given a choice of colours to choose from for the background. Whenever the participant appeared distracted due to the 2D animation, the interface was used without the 2D animation, or the size of the window was controlled, depending on the priority of need.

T-test was conducted on the two sets of data (group interface and with group interface). The t-test shows that the two sets of data are significantly different at the 1% level (0.00014). Hence it can be deduced that using the personalised interface is 99% better than using the group interface.
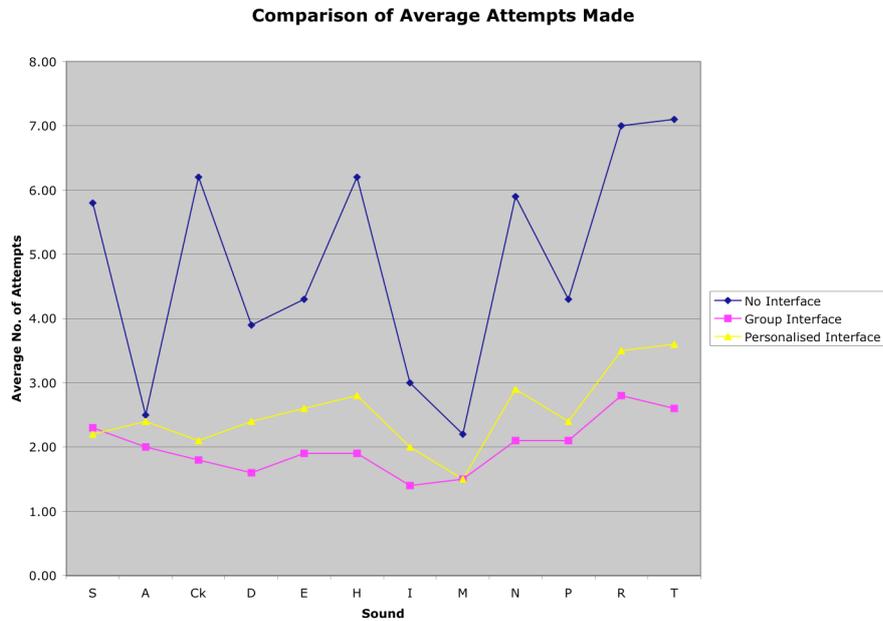
Comparison of Average Attempts Made

Fig. 4. Comparison of the three experimental results

The group and personalised interfaces have a clear advantage over the no interface data. However the personalised interface goes on to improve the pronunciation even further as indicated in the graphs above. Participants 2 and 9 were also able to pronounce the sounds without the audio indicating that the 3D component of the interface was accurately designed to indicate the actual movements of the mouth.

The qualitative and quantitative evaluations of the developed interfaces indicated that they had a clear advantage over having no visual representation and personalised indicated that the participants of the personalised interface performed with a better success rate than the group interface as indicated by t-test and the graph.

6   Conclusion and Future

This research went on to design, develop and evaluate a novel multimedia interface, which can be used as an assistive device to enhance their pronunciation skills so that they can be educated in the mainstream schools. Even if solutions are found for mental effects biological factors such as having a larger tongue in proportion to the mouth (e.g. Down's Syndrome), may not change. Therefore, there will be a continuous need for further research in this area.

This interface could be adapted easily for other language phonetical training. Work is also done to further improve the interface to include deaf children.

There have been many attempts to create an interface to enhance pronunciation none has been completed successfully. This fact was emphasised by the speech therapist who participated in the focus group which supported the design and development process of this multimedia interface. This work has been very much a pioneering work and has achieved the target it pursued. Further research is expected to be done to move this milestone even further.

References

1. Reuters., (2006), "Mps blast special needs system", http://uk.news.yahoo.com/06072006/325/mps-blast-special-needs-system.html, Accessed 6th July 2006.
2. Buckley, S., (1999), *Improving the speech and language skills of children and teenagers with Down syndrome*, Down Syndrome news and Update Vol.1, No. 3, pp 111-128, The Down Syndrome Educational Trust, Great Britain.
3. Bowen, C., (2005), PsychNet-UK, http://members.tripod.com/Caroline_Bowen/phonol-and-artic.htm, Accesses on 17th March 2006.
4. Jolly Learning (2006), http://www.jollylearning.co.uk/jp.htm, Accessed 27th July 2006
5. Bowen, C., (2001), http://meembers.tripod.com/~Caroline_Bowen/audbom.html
6. Stone, D, Jarrett, C, Woodroffe, M, Minocha, S (2005), *User Interface Design and Evaluation,* The Open University, UK: Elsevier, Morgan Kaufmann Publishers.
7. Gimson, A.C., (1980), *An Introduction to the Pronunciation of English,* Edward Arnold Publishers, UK.
8. AbilityHub, (2006), http://www.abilityhub.com/aac/aac-devices/htm, Accessed on 13th March 2006.
9. Sevcik, R.A., Romski, M.A., Adamson, L.B., (2004), *Research directions in augmentative and alternative communication for preschool children*, Disability and Rehabilitation, Vol. 26, No. 21/22. 1323-1329, Taylor & Francis Ltd.
10. Gong, L., (2003), *Human and Humanoid don't' match: Consistency preference and impact on users' trust,* Interact 2003, IOS Press, Zurich, Switzerland (160-167).
11. Meyer, J., (1998), *Teach me to talk,* http://www.csun.edu/cod/conf/1998/proceedings/csun98_039.htm, Accessed 17th July 2006.
12. Watanabe, T., Okubo, M., Danbara, R., (2003), *InterActor for human interaction and communication support,* Interact 2003, IOS Press, Zurich, Switzerland (113-120).
13. Bälter, O., Engwall, O., Öster, A., Kjellström, H., (2005), *Wizard-of-Oz test of ARTUR – A computer-based speech training system with articulation correction,* ASSETS 2005, October 9-12, Baltimore, Maryland, USA, ACM.
14. Lukaszewicz, K., (2003), *The ultrasound image of the tongue surface as input for man/machine interface,* Interact 2003, IOS Press, Zurich, Switzerland (825-828).