Using Multimedia Based Assistive Technology to Enhance Pronunciation Skills for Children with Phonological Disorders

Jennifer George[1] and Paul Gnanayutham[2]

[1]SAE Institute, United House, North Road, London, N7 9DP, United Kingdom
[2]Department of Computing, University of Portsmouth, Buckingham Building, Lion Terrace, Portsmouth PO1 3HE, United Kingdom
{jennifer.george@sae.edu, paul.gnanayutham@port.ac.uk}

Abstract. Children are often forced into mainstream schools while special needs schools are being shut down. Children with phonological disorders in mainstream schools go through fear, shame and frustration of not being understood. The proposed research attempts to address this issue by way of designing and developing a prototype for an assistive device that would help special needs children in mainstream education. This will also help children overcome problems that result from lack of confidence. This work has been very much a pioneering work and has achieved the target it pursued. This study was carried out in two phases first one with a group interface and the second phase with a personalised interface that catered for the needs of each individual user.

Keywords. Phonological disorders, speech disability, pronunciation, interface, sounds and children.

1   Introduction

This section introduces the basics of a sound is produced. The seven organs of speech are tongue, teeth, lips, muscles of the lower jaw, hard palette, soft palette and teeth ridge (also known as the alveolar ridge). 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. Sander (1972) created a Speech Sound Development Chart identifying the range of typical sound development in children according to their age. When unusual patterns of this developmental process occur in children the advice of Speech and Language Therapists (SLT) is contacted. These unusual patterns may be diagnosed as either disordered or simply delayed (Lemon and Bucke 2006). Disordered speech may occur due to poor auditory discrimination, difficulty in forming sounds, structural difficulties, verbal dyspraxia or hearing difficulties (Bowen 2005). These disorders on its own can be mild, moderate or severe and connected to other disorders. It may also be as a result of basic communication disorders. The child's chronological age and developmental level may differ according to the mental and biological growth of each child. Buckley (1999), states that the main difficulties experienced by children with Downs Syndrome can be categorised as difficulties in hearing, auditory perception and processing, difficulties in clear speech production and greater difficulty in learning grammar than vocabulary. Children with learning disabilities need more high quality

learning opportunities in order to learn and remember the forming of sounds, yet they get much less opportunities because of their slower progress.

This tool used multimedia to create an application that any support worker or parent will be able to use with minimum training with the children to enhance communication with their peers. Having established the need for a device to enhance pronunciation, the research problem was "*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?*" explored. The investigation was carried out using still images, two-dimensional and three-dimensional animations and evaluated with a focus group and individual participants with speech impairments.

2   Assistive Devices

Many multimedia interfaces have been developed as both Augmentative Alternative Communication (AAC) and Assistive Technology (AT) to aid speech disabilities (AbilityHub 2006). Children with severe communication impairments use alternative communication modes such as manual signs and visual symbols, which is a substitution for spoken words (Sevcik *et al.* 2004). 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. This research will only be dealing with AT as this encourages speech and does not attempt to replace it with alternative methods.

3   Chosen Methodology

The design, development and experiments were carried out in two cycles and two phases (Fig. 1.). The two cycles were iterative processes that made the final versions of the group and personalised interfaces possible prior to conducting the experiments. Phase one experiment was conducted on the general interface to explore whether a group of interaction paradigms will enhance the performance of pronunciation skills for children with speech impairments and phase two experiments were conducted with a personalised interfaces to explore whether one novel interaction paradigm can be utlised to enhance the performance of pronunciation skills for children with speech impairments.

4   Design and Development

A focus group shown in Table 1, centred around two disabled and eight able was used to evaluate the interfaces during the design and development process. Changes were made to the interface in an iterative process. The interface to be one that can be used on a PC or MAC platform with a standard CD-DVD Rom, speakers and Flash Player (Fig. 2 and 3). The relevant audio clips for the animations were recorded using Pro Tools. They were recorded in both male and female voices and lasted for almost three seconds. The information as to how each sound was produced was obtained from the teaching material used by the Wendy Whatmore Academy of Speech and Drama. Jolly Phonics is a speech sound training system used by many therapists as derived from the various interviews (Jolly Learning 2006). Sounds in order of frequency of use were selected for prototype and the relevant actions for these sounds were obtained from the Jolly Phonics online resources for the 2D animation.

Maxon's Cinema 4D was used due to the simplicity and the ease of bone animation. This was exported as a .mov file so that evaluation could be conducted online and offline. Macromedia Flash was chosen to create the main part of the interface. The interface could be exported to a CD or DVD and used for evaluation. The user should be able to quit the interface at any time and for this purpose, an exit button has been included. Volume could be controlled using the hardware's volume control. Relevant 2D animation to demonstrate an example of the sound is included to help the child relate to the sound. Each 3D Animation was included with the relevant sound produced to demonstrate how the organs of speech are used in formation to create sound. The 2D and the 3D animations were synchronised with the relevant audio. There was no need for a pause button as the animation would not last for more than 3-5 seconds. The start and stop buttons could be replaced with a replay button. The 3D window had to be significantly larger than the 2D window.

The Jolly Phonics actions for sounds would be used to design and develop the example animations (Jolly Learning 2006).This would be exported as .swf and thereafter imported into the authoring interface. The standard system of learning sounds used by the NHS is Jolly Phonics, which consists of the 42 sounds in English. This method divides the sounds into seven groups. The first group of six sounds have been chosen according to the frequency in three-letter words compared to the second group. Similarly the less used sounds come later on in the list. When these two groups are compared to the International Phonetic Alphabet (IPA) it can be concluded that all organs of speech are being used during the learning of these twelve sounds (Ager 2006). This list has a good mixture of both consonants and the high frequency vowels. Thus, the first twelve sound. Sounds from the Jolly Phonics System s, a, t, i, p, n, ck, e, h, r, m and d were chosen. Jolly Phonics also gives actions for each sound to help the children remember letters that represent it. This idea was adopted in order to give the child a demonstration in the form of 2D animation to help the child remember and associate with the sound produced (Jolly Learning 2006).

Semi-transparent human 3D models have been used to demonstrate the formation of different sounds inside the mouth. The head and shoulder of a boy and a girl with well defined parts inside the mouth were modelled. The lips, teeth, teeth ridge (alveolar ridge), tongue, hard palette, lower jaw, soft palette, Velar (back of the tongue and near the soft palette), uvula (extension at the back of soft palette) are visible.

Modelling, texture and lighting, animation and camera movement and rendering of clips using the audio files of the twelve chosen sounds are to be created in C4D. The animations are based on the organs used in formation for each of these sound as defined below. Initially, the mouth can be animated and thereafter they can be exported separately for both the boy and the girl. Each animation may last up to three seconds showing clearly the firm contact and release of the relevant organs of speech in formation.

Gimson (1962) described the production of each sound in the English language. This was used to create animation in order to distinguish, explain the formation and model the creation of each of these sounds. The letters used in the interface depict the sound that is produced and not the alphabet. The animations would always start from a relaxed

position i.e. Tongue lying flat and upper teeth over lower teeth and following animation regain this position.

Various shapes were considered for the shapes of buttons and the interface objects (Moran and Leiser 2002). Rectangle was chosen having researched into Gestalt Laws and psychology of visual perception (Ware 2000). Previous investigations show that users have emotional reactions to colours and fonts, this interface gave the option for making these changes to suit any user. This study also showed that white or yellow text on blue background was more readable (Laarni 2003) and this was taken as the default setting for the interface while lilac is to be the background colour for the individual animations. The buttons are to be yellow turning into pink on rollover. The active sound would also be a lighter yellow. The 3D animation colours are natural and do not change.

Still images will be used to display the front view of the mouth in 2D form. There will be two moving images, one demonstrating the sound made using side view and the other indicating an example of the sound made. This will illustrate movement of the mouth necessary to produce the sound.

Two types of audio were used for the interface. An introductory sentence with instruction to click on buttons and a thank you on exit has been used. For each demonstration of sound, the sound produced is also played.

The final version of the group interface was taken as the starting point for the personalised interface. This interface was used for experiments carried out with participants during which time participant observation and action research were undertaken.

In this section, the theory of widgets was used to design the preferences page to give personalisation functionality to the interface. Layered approach was used to design the preferences functionality of the interface. Layout changes for the main page also had to be made occasionally to accommodate the personalisation. Options were given in choosing between male and female voices, switching the 2D and 3D animations off and a choice of window size adhering to usability issues.

To cater to the visually impaired and different types of disability with different levels of distractivity options in changing the window size and turning in completely off were given. The instructor's voice was also an option to choose from a male and female voice. Although this interface had various properties to cater for different age groups, an option to minimise or exit the window which was distracting the user e.g. An older child may not need the 2D animation for examples.

Following the development of the final interface of both group and personal, experiments were carried out to test the research question. This interface was used for experiments carried out with participants during which time participant observation and action research were undertaken.

5   Experiments and Results

Phase one of the research lasted 6 weeks and produced the final general interface. Phase two of the research lasted 8 weeks and produced the final personalised interface. This research created the personalised interface effectively by conducting extensive experiments with a focus group before using the interface with the disabled participants. First experiment was carried out following the low fidelity prototype development

evaluation, experiment two was carried out during phase one and experiment three was carried out during phase two of this investigation. Statistical analysis was carried out and conclusions were reached. Table 2 gives details of the participants who evaluated the group interface. They were between the ages of 5 and 36 years, 4 females and 6 males, 7 able and 3 disabled.

Exploratory Phase:
This phase of the study was carried out to investigate:
1. The number of attempts needed to pronounce each sound;
2. The amount of training needed;
3. The sounds that cause difficulty.

The participants were asked to repeat the 12 sounds after the researcher (Table 3). Each sound had to be pronounced and the anatomy explained. A maximum of ten attempts were given for each sound by the participant. Thereafter they were indicated as '10'. Average attempts made by the participants to pronounce each sound correctly were calculated (Table 4).
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.

Phase One
This phase of the study was carried out to investigate whether:
1. Disabled participants can be grouped together, when developing interfaces;
2. Users can use Group Interface with minimal training;
3. Optimum setting for group interface can be obtained with focus group, which can be used as default when evaluating the interface with disabled users.

Apparatus used was, Computer with CD-DVD driver and speakers, interface program. 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 (Table 5). Statistical analysis was conducted from the data obtained to choose the optimum initial settings for the interface. Average of attempts made were calculated, in order to compare profiles.
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. In this phase of the research, participants were grouped and experiments were conducted to evaluate the group interface with minimum training. Results were recorded; data analysed and statistical analyses were carried out. Outcomes are to be compared with the results from the personalised interface in the next phase of the research.

Phase Two

This study investigated whether:

1. Using the Personalised Interface improved the performance of the disabled user in comparison with the group interface;
2. Personalised Interface can be reconfigured anytime;
3. Users can use Personalised Interface with minimal training;
4. Optimum setting for Personalised Interface can be obtained with able focus group, which can be used as a starting point when evaluating the interface with the participants.

The functionality of the interface was explained to the participant by way of demonstration. Whenever necessary it had to be explained to the participant that this was what the inside of the mouth looked like. The participant was asked to imitate the 3D animation for each sound. Maximum of ten attempts were given for each sound. 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. Statistical analysis was conducted from the data obtained to choose the optimum initial settings for the interface. Similar to experiments one and two, the average attempts made were calculated, in order to compare profiles.

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.

In this phase of the research, participants individually evaluated the interface and experiments were conducted to evaluate the personalised interface with minimum training. Results were recorded; data analysed and statistical analyses were carried out. Outcomes were compared with the results from the group interface in the previous phase of the research. The comparison in figure 4 indicated that the performance of the personalised interface was much better than the group interface.

Figure 7 shows the number of attempts made to pronounce each sound using the various interfaces. The graphs shows the performance of each interface supporting the t-test results calculated earlier. The group and personalised interfaces have a clear advantage over the no interface data. However the general interface goes on to improve the pronunciation even further as indicated in the graphs Figure 4. 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, which indicated that the interface could be used with participants with hearing difficulties.

6   Conclusion

This research carried out a survey of AAC tools and AT devices to find out whether there were any devices that could be used by children with speech impairment to enhance their pronunciation skills so that they can communicate with confidence in main stream schools without trailing other students in a class. This need arose since the schools for special needs were being closed down and children with special needs were made to join the main stream education. The survey indicated the lack of an assistive device to enhance the pronunciation skills for this group of children with phonological disorders, hence the motivation for this research.

There were challenges involved with developing and evaluating interfaces for children with phonological disorders, approaches chosen for this research, ethical issues involved and the structure of the investigation. Various methods and methodologies were considered and appropriate methods and methodologies were chosen for the research while others were 'excluded by including'. Having established the need for a multimedia interface to enhance pronunciation, the research problem on the type of interface needed was answered successfully.

Two interfaces were developed to answer the research question. The developed multimedia interface grouped the sounds according to the mechanism in which they are produced in the mouth and also had options in choosing the size and colour of the window and the trainer's voice. The developed interfaces were evaluated by a focus group of ten participants. The qualitative and quantitative evaluations of the developed interfaces for group and personalised indicated that the participants of the personalised interface performed with a better success rate than the group interface as indicated by the t-test and the graph.

Even if new solutions were found for mental effects biological factors such as 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. Some indication of the future research is dealt in the next section.

7   Future Plans

Pattern recognition languages may assist user centred design process (Maybury 2003, Mahemoff and Johnston 1998). One way to ensure usability is to treat human abilities as input to design rather than merely evaluating prototypes. Freeman (1996) proposed a theory of brain dynamics to construct external representations by observing brain patterns. He stated that minds were nothing but activity of neurons and that thought and ideas were processed according to rules, discovered by psychologists. With an external input device for brain waves this interface can be used for the brain injured during speech therapy. Gnanayutham and his team (2004) designed and developed a soft keyboard for non-verbal quadriplegic brain injured persons. Future research in this area could connect brain waves to the interface developed in this research thereby taking this research to the next stage where the brain injured non-verbal community could be integrates into this research (Gnanayutham *et al.* 2005).

8   Bibliography

AbilityHub, (2006), http://www.abilityhub.com/aac/aac-devices/htm, Accessed on 13th March 2006.

Ager, S., (2006), http://www.omniglot.com/writing/ipa.htm, Omniglot, Accessed 29th August 2006.

Bowen, C., (2005), PsychNet-UK, http://members.tripod.com/Caroline_Bowen/phonol-and-artic.htm, Accesses on 17th March 2006.

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.

Freeman, W. J., (1996), *Brain dynamics in the genesis of trust as the basis for communication by representations,* AAAI 1996, MIT Press (1327-1328).

Gimson, A.C., (1962), *An Introduction to the Pronunciation of English*, Edward Arnold Publications.

Gnanayutham, P., Bloor, C., Cockton, G., (2004), Soft-Keyboard for the disabled, Miesenberger, K., Klaus, J., Zagler, W., Burger, D., (Eds.), ICCHP2004, July 2004, Springer-Verlag Publishers, Paris, 999 - 1002.

Gnanayutham, P., Bloor, C., Cockton, G., (2005), *Discrete Acceleration and Personal Titling as Brain-Body Interface Paradigms for Neurorehabilitation,* CHI2005, April 2005, Oregon, USA.

Jolly Learning, (2006), http://www.jollylearning.co.uk/jp.htm, Accessed 27th July 2006

Laarni, J., (2003), Effects of color, font type and font style on user preferences, Edited by Stephanidis, C., Adjunct Proceedings of HCI International 2003, June 2003, Crete University Press, Crete, 31 – 32.

Lemon, K., and Bucke, L., (2006), Interviewed 18th July 2006

Mahemoff, M.J., Johnston, L.T., (1998) *Principles for a usability-Oriented Pattern Language,* Ozchi, pg.132, Australian Computer Human Interaction Conference, 1998.

Maybury, M., (2003), *Intelligent user interfaces: An Introduction,* Tutorial presented at MITRE Corporation, http://www.1uiconf.org/03tutorial1.html.

Moran, S., Leiser D., (2002), *The limits of shape constancy: point-to-point mapping of perspective projections of flat figures,* Behaviour & Information Technology, 21(2), 97 - 104

Sander, E. K., (1972),
http://stcloud.k12.mn.us/ecassessment/assets/SpeechSoundDevelopmentChart.pdf#search=%22When%20are%20speech%20sounds%20learned%20-%20Sander%2C%20Eric%20K%22, Accessed 29th August 2006.

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.

Ware, C., (2000), Information Visualization, Morgan Kaufman Publishers, USA, 203 – 213.

Table 1 - Focus Group

Capacity	Gender	Age	Institute
Developer of interfaces for brain injured	M	51	University of Portsmouth
Sibling/Carer of adult with Down Syndrome	F	30	Somerset
Speech Therapy Consultant	F	64	Redhill, Speech and Language Therapist
Assistant Speech Therapist	F	29	Radcliffe Infirmary, Oxford
Speech Therapist	F	42	Ormerods, Oxford
Speech Therapist	F	38	Oxford City PCT
Adult with Down Syndrome	F	32	Somerset
Multimedia Developer	M	29	SAE Institute, London
Multimedia Developer	M	31	SAE Institute, London
Child Down Syndrome	M	10	Croydon

Table 2 - Details of the participants used in experiments

Part. No	Institute	Age	Gender	Clinical Diagnosis
1	Home	7	Male	None
2	Home	4	Male	None
3	Home	5	Female	None
4	Home	6	Female	None
5	Home	7	Male	None
6	Home	9	Male	Autism
7	Home	13	Male	Downs Syndrome
8	Home	6	Female	None
9	Home	5	Male	None
10	Home	36	Female	Downs Syndrome

Table 3 - Results of experiment without interface

Participant	Number of Attempts per Sound

Participant	1	2	3	4	5	6	7	8	9	10
s	5	6	5	4	7	5	4	6	7	9
a	3	2	1	2	2	6	3	3	2	1
ck	5	6	7	5	5	10	10	4	8	2
d	3	4	5	4	2	6	7	3	4	1
e	6	4	3	4	7	3	4	7	2	3
h	7	5	4	3	6	10	9	6	4	8
I	2	1	4	2	3	3	2	3	4	6
m	2	1	2	2	2	2	3	2	3	3
n	6	5	6	5	4	9	8	5	6	5
p	3	4	5	4	3	7	5	4	3	5
r	3	8	9	7	3	10	10	3	7	10
t	7	6	5	5	8	9	10	8	7	6

Table 4 - Sucess Rate without Interface

Sound	Average Attempts Made

s	5.80
a	2.50
ck	6.20
d	3.90
e	4.30
h	6.20
i	3.00
m	2.20
n	5.90
p	4.30
r	7.00
t	7.10

Table 5 - Results for General interface

Participant	Number of Attempts per Sound									
	1	2	3	4	5	6	7	8	9	10
s	2	3	3	3	2	5	2	2	2	pending
a	2	3	3	3	2	4	2	2	2	
ck	4	3	3	3	2	4	3	2	2	
d	2	2	4	4	2	3	4	2	2	
e	2	4	3	3	2	3	4	4	2	
h	2	3	3	3	2	5	3	2	2	
I	2	2	3	3	2	2	3	2	2	
m	2	2	3	3	2	2	3	2	2	
n	2	2	3	2	2	6	3	4	2	
p	3	3	4	3	3	3	4	4	3	
r	2	2	3	3	2	7	5	3	2	
t	3	3	4	4	3	4	4	4	2	

Table 6 - Success Rate for General Interface

Sound	Average Attempts Made
s	2.30
a	2.00
ck	1.80
d	1.60
e	1.90
h	1.90
i	1.40
m	1.50
n	2.10
p	2.10
r	2.80
t	2.60

Table 7 - Results for personalised interface

Participant	Number of Attempts per Sound									
	1	2	3	4	5	6	7	8	9	10
s	2	2	3	2	1	3	2	2	2	3
a	2	3	3	3	2	4	2	2	2	1
ck	3	2	3	3	1	4	1	1	1	2
d	2	2	4	2	2	3	4	2	2	1
e	2	2	3	3	2	3	4	4	1	2
h	2	3	3	3	2	5	3	2	2	3
I	2	2	3	2	1	2	3	2	2	1
m	1	2	2	3	1	1	1	1	2	1
n	2	2	3	2	2	6	3	4	2	3
p	1	3	2	3	3	3	3	1	2	3
r	2	2	3	3	2	7	5	3	2	6
t	3	3	4	4	3	3	4	4	2	6

Table 8 - Success Rate of Personalised Interface

Sound	Average Attempts Made
s	2.20
a	2.40
ck	2.10
d	2.40
e	2.60
h	2.80
i	2.00
m	1.50
n	2.90
p	2.40
r	3.50
t	3.60

Fig. 1 – Chosen Methodology
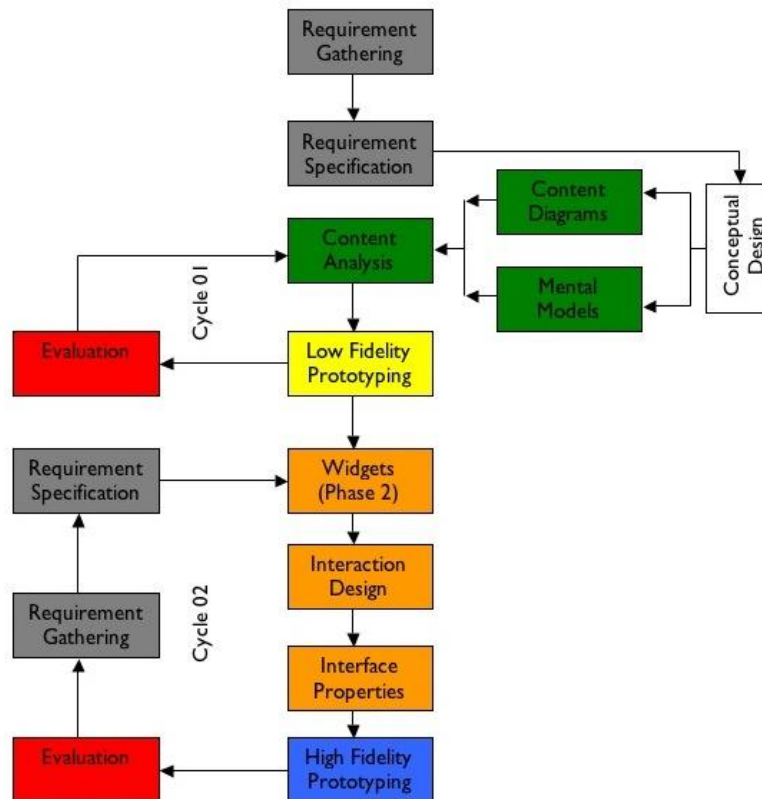
Fig. 2 – Developed Interface
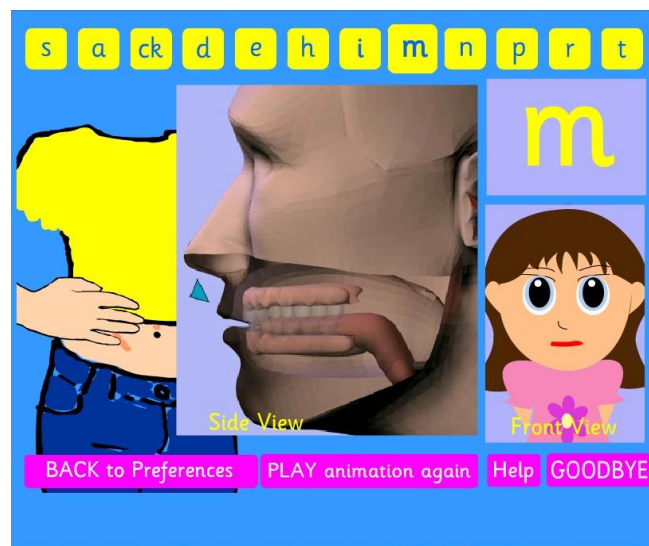
Fig. 3 – Preferences to personalise interface

Fig. 4. - Comparison of Results from the three phases