

Developing Multimedia Interfaces for Speech Therapy

¹Jennifer George and ²Paul Gnanayutham

¹*Department of Computing, University of Portsmouth, Buckingham Building, Lion Terrace, Portsmouth, PO1 3HE, United Kingdom* ²*SAE Institute, United House, North Road, London N7 9DP United Kingdom*

¹+44 (0)207 609 2653, ²+44(0)23 9284 6468

¹+44 (0)207 609 6944, ²+44 (0)23 9284 6364

jennifer.george@sae.edu, paul.gnanayutham@port.ac.uk

www.jennifer-george.co.uk, www.paulgnanayutham.com

Abstract. Multimedia has been used creatively to entertain and educate, it can also be used for therapeutic and medical purposes. This research addressed this issue by incorporating multimedia to design and develop an assistive device that would help special needs children in mainstream education. The appropriate methodology for developing such an interface was looked into. The relevant multimedia, psychology, social and educational theories were taken into account. Based on this literature, interfaces to enhance pronunciation were designed, developed and tested.

Keywords

Phonological disorders, HCI, Inclusive design, Assistive Technology

1 Introduction

Reuters reported that children are often forced into mainstream schools while special needs schools are being shut down [1]. Literature survey indicated the need for an assistive device for this group of children who had to study in the main stream of education but with a limited support for their disability. Research was carried out to answer the question whether such as device will enhance their pronunciation skills and what should be the properties of such an interface. Having established the possibility of a need for an assistive device, there arose a research problem on the type of

interface needed. The research problem to be addressed was:

“Should there be a group interface or one interface that can be personalised, to enhance the performance of pronunciation skills for children with speech impairments?”

Technology-related assistance for individuals with disabilities Act of 1988 (TECH Act) was designed to enhance the availability and quality of Assistive Technology (AT) and services to all individuals and their families throughout the U.S. This Act defines AT devices as an item, piece of product or a product system that is used to increase, maintain or improve functional capabilities of individuals with disabilities. Examples of Cognitive AT devices include multimedia CD-Rom based application programs and CD-Books.

This article reports on the design and development of an artefact using multimedia to address the need for an assistive device to rehabilitate children with phonological disorders using traditional therapeutic methods.

2 Methodology

This investigation was carried out in two phases. Permission was obtained from parents/guardians or carers of each participant used in these experiments. A focus group of ten able bodied participants, consisting of speech therapists, carers, parents and multimedia designers were used for

evaluation throughout this iterative development. The first phase was used to develop an interface that can be used for a group of disabled children while the second phase was used for developing an interface which could be personalised to cater for each child on an individual basis. The experiment involved getting the children to imitate the sound displayed by animations. Phase two included similar evaluation and experiments but with more options to be able to personalise the interface. Formative and summative evaluation was carried out with participants of phase one and two to obtain

optimum data for accuracy of information and usability of multimedia. Results obtained during the two phases were both recorded and analysed.

The methodologies and methods used are arranged as per figure 1 [2].

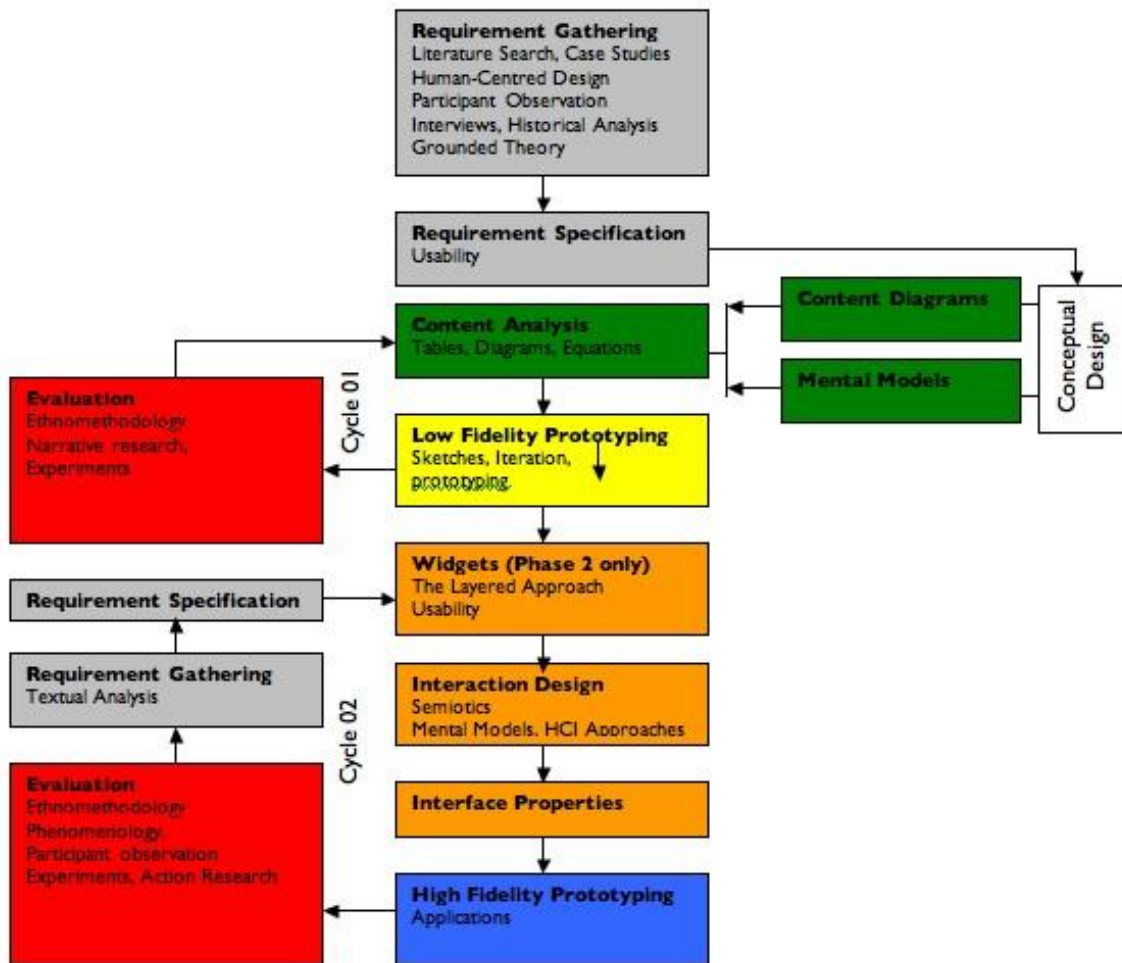


Figure 1 – Research Methodology

3 Theories

Multimedia is a form of human communication that is computer-based, it combines media and is interactive. Many multimedia practitioners have derived at a number of theories based on traditions of literacy, philosophy and media

theory. For a multimedia device to be designed and developed existing theories on related areas such as multimedia, psychology, education, society and speech were studied and appropriate theories were chosen for the research while others were ‘excluded by including’. These theories are included in various stages of the research as per diagram below. Multimedia is a form of human communication that is computer-

based, it combines media and is interactive. Many multimedia practitioners have derived at a number of theories based on traditions of literacy, philosophy and media theory. These theories can be used as guidelines for interface designs. Educational theories can be utilized in the planning and delivery of content. “Theories in psychology are powerful instruments, admirable constructs, and they satisfy important needs to explain and understand ourselves, our

interactions with others and with our environment” [3]. This section covers the various psychological theories related to the research topic. Social theory refers to systematic study of relationships and behaviour. This includes a number of modern and post modern theories. Some of these theories apply at various stages of this research and have been arranged according to the research methodology in figure 2 [4].

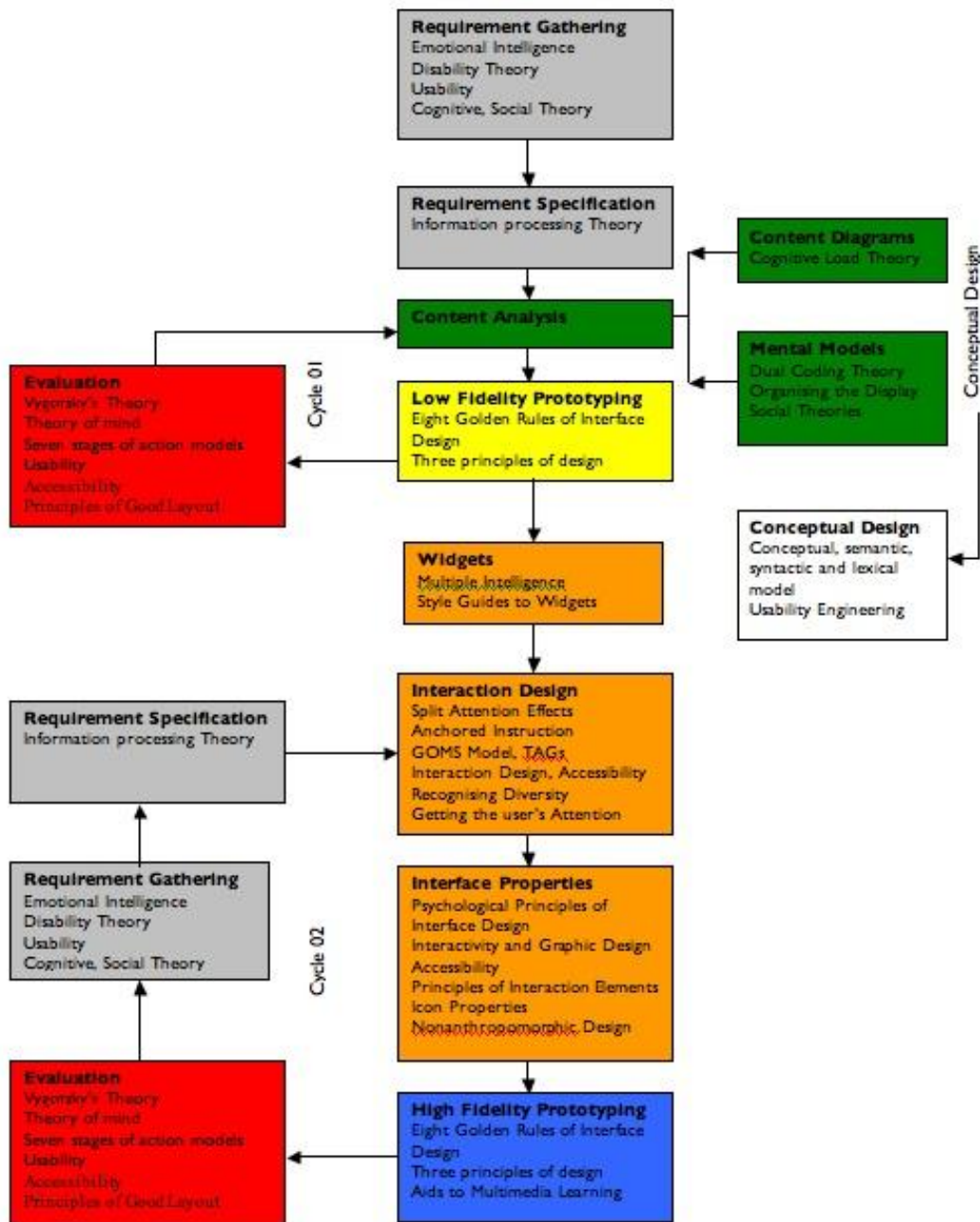


Figure 2 – Applied theories

4 Design and Development

4.1 Requirement Gathering

Literature Search was carried out to investigate the formation of sounds, speech disorders, therapeutic methods, existing interfaces and relevant theories and methods used for topics similar to the research title. Case study was

carried out from the reports of the Sara Duffen Centre for the Downs Syndrome in Portsmouth, UK. This process was centred around two disabled and eight able participants who functioned as the focus group (table 1). Participant Observation was undertaken in two institutions in Oxford and they were included in the Human Centred Design process. Initial interviews were carried out with the focus group. Historical Analysis was undertaken at two institutions in Oxford and an institution in Redhill. Grounded Theory was used to establish the need for the research.

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, Home
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, Speech and Language Therapist
Adult with Down Syndrome	F	32	Somerset, Home
Multimedia Developer	M	29	SAE Institute, London
Multimedia Developer	M	31	SAE Institute, London
Child Down Syndrome	M	10	Croydon, Home

Table 1 – Focus Group

4.2 Requirement Specification

The proposed interface would have an easy navigation, a 3D demonstration of sound created with audio and a 2D animated example of the sound. The interface should be one that would have both visual and auditory output.

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 Gimson [5]. Jolly Phonics is a speech sound training system used by many therapists as derived from the various interviews. 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.

Bryce, Maxon's Cinema 4D and Maya were considered for creation of the 3D property of the interface. Maxon's Cinema 4D was decided due

to the simplicity at the same time the functionality of the requirement. This would be exported as a .mov file. Macromedia Flash and Macromedia Director were considered to be the authoring tool. As evaluation may also be conducted online and the online support available Macromedia Flash was chosen to create the main part of the interface. The components of the interface to be included in the low fidelity prototype followed by the high fidelity prototype are as follows:

4.2.1 Controls

On the main interface, facility to start again should be provided as this could be used for reinforcement. The user should be able to quit the interface at any time and for this purpose, an exit button should be 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 are synchronised with the relevant audio.

4.2.2 2D Objects

The Jolly Phonics actions for sounds would be used to design and develop the example animations [6]. This would be exported as .swf and thereafter imported into the authoring interface.

4.2.3 3D Objects

Purpose : Demonstrating using semi-transparent human models the formation of different sounds inside the mouth.

Models : Head and Shoulder of a boy and a girl with well defined parts inside the mouth.

Parts of the mouth : Lips, teeth, teeth ridge (alveolar ridge), tongue, hard palate, lower Jaw, soft palate, Velar (back of the tongue and near the soft palate), Uvula (extension at the back of soft palate)

4.3.4 3D Animation

Modeling, 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 the 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.

4.3 Conceptual Design

Content diagrams were created in order to analyse the facilities needed in the interface (Figure 3). Mental models were made in order to arrange the link between the different pages/scenes and the arrangement of objects within each page (Figure 4). Sketches were drawn using tree diagrams, menus and storyboards.

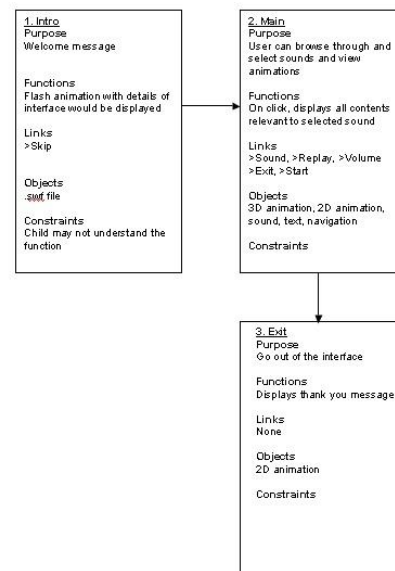


Figure 3 – Sample content diagram

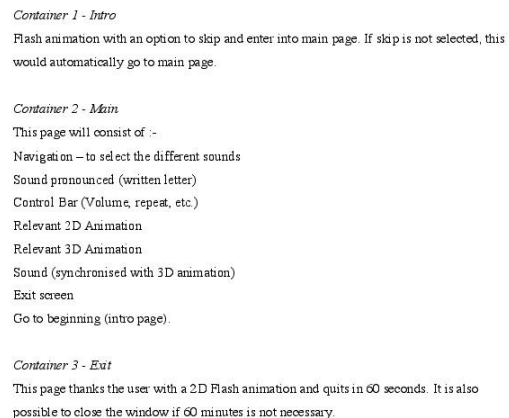


Figure 4 – Sample mental model

4.4 Content Analysis

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 [7][8]. This list has a good mixture of both consonants and the high frequency vowels. Thus, the first twelve sounds from the Jolly Phonics System [6] 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 [6].

S - Weave hand in an *s* shape, like a snake, and say *sssss*

A - Wiggle fingers above elbow as if ants crawling on you and say *a, a, a*.

T - Turn head from side to side as if watching tennis and say *t, t, t*.

I - Pretend to be a mouse by wriggling fingers at end of nose and squeak *i, i, i*.

P - Pretend to puff out candles and say *p, p, p*.

N - Make a noise, as if you are a plane - hold arms out and say *nnnnnn*.

CK - Raise hands and snap fingers as if playing castanets and say *ck, ck, ck*.

E - Pretend to tap an egg on the side of a pan and crack it into the pan, saying *eh, eh, eh*.

H - Hold hand in front of mouth panting as if you are out of breath and say *h, h, h*.

R - Pretend to be a puppy holding a piece of rag, shaking head from side to side, and say *rrrrr*.

M - Rub tummy as if seeing tasty food and say *mmmmmm*.

D - Beat hands up and down as if playing a drum and say *d, d, d*.

The International Phonetic Alphabet (IPA) explains the formation of each sound specifically mentioning the organs of speech used in formation [7].

The letters below 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. What happens in between is described below.

S – lift tongue slightly lifted in the tip to be position between the upper teeth

A – drop jaw to open mouth (wider than sound ‘e’) and bring tongue towards the lower lip

T – as you open mouth slightly, lift tip of tongue between the hard palate and the teeth ridge and release

I – open mouth slightly and stretch lips wide and show teeth

P – close lips firmly, push toward inside of the mouth and release

N – lift tongue up and press on the hard palate and release (see if you can show something happening in the nose)

CK – press the muscles in the throat together (towards the centre of the throat) and release

E – smile while having mouth slightly open (tongue does not move)

H – open mouth slightly (don’t smile as with as ‘e’) (and release air from throat)

M – press both lips together and release (see if you can show something happening in the nose similar to ‘n’)

D – lift middle of tongue, touch hard palate and release

4.5 Low fidelity Prototyping

Storyboarding is carried out to define the sequence of a story keeping in mind, the resources available and the proposed time. This is used to organise the focus of a story. This technique is widely used in the world of multimedia. Draheim and Weber researched in two staged interactions, a concept behind form based interfaces [9]. From this we obtain storyboarding a method for eliciting, specifying and communicating functions for form based interfaces. This method comprises a visual language and a set of proposals for activities. The visual language is based on bipartite state transition diagrams. The HCI committee believed that storyboarding will be replaced by some advanced interface but there seemed to be no interface metaphors able to provide a real

alternative. Form based interfaces are self-explanatory and intuitive.

Based on the content diagrams and mental models created, the storyboards and sketches were created (figure 5).

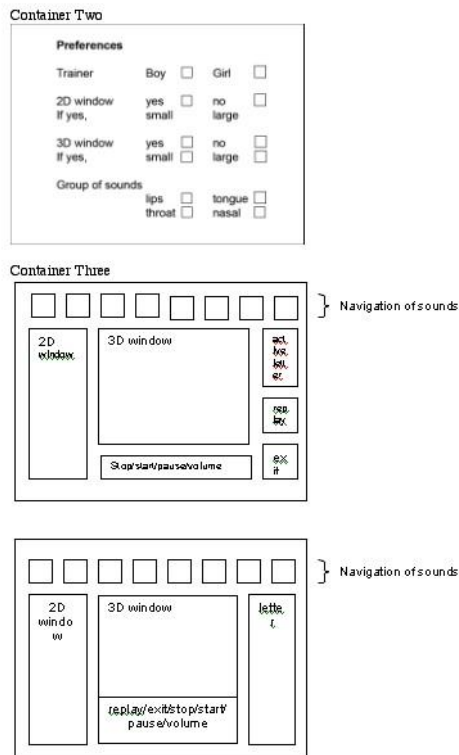


Figure 5 - Storyboards

4.6 Evaluation

Ethnomethodology was used to gather feedback from focus group. Narrative research was carried out on cultural and educational backgrounds of participants that may influence perception. The following sketch was agreed upon for the design of the main page. It was agreed that the content analysis was feasible and could be implemented effectively. The conclusions of the focus groups were as follows. The first sketch for container three had too many windows and there was no need for volume control as this could be done using the hardware itself. The second sketch for container three had too much space for the letter and had too much functionality within the navigation. 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 low fidelity prototype was further developed in order to start the high fidelity prototyping.

4.7 Interaction design

Semiotics, Mental models and HCI principles were used in defining the interactivity of the interface. Macromedia Flash was chosen for the main interaction design. Depending on the size requirement of the preferences, Flash files could be scaled. Flash files could also be exported independent of operating systems and would remain consistent regardless of the platform available with free downloadable plug-in. Integration of multimedia files is easy compared to other application while containing original files and fonts. The file sizes are small as all files are contained within a single flash file nevertheless containing high quality images. Flash's disadvantages were also considered. If published online, it would require an obligatory plug-in. Search engines can hardly find Flash and Flash sites cannot be bookmarked. Advanced functionality needs experience in Flash's dedicated scripting language Actionscript. The output is a divergence from known objects from the Internet.

Holzinger and Ebner state that when designing and developing animations two essential factors to be considered would be interaction design and graphic design [10]. Both these aspects are addressed using different authoring tools especially vector based Flash technology. Flash encourages superfluous animation that can cause poor design. It is essential that HCI theories are considered while enhancing the design.

Berglund and Qvarfordt used speech input to augment the remote control interaction techniques for television [11]. They explored possible error resolution strategies in speech recognition systems from using techniques such as repetition of input by the user and choice of interpretation. These two techniques were tested using Wizard-of-Oz method wherein the user is observed unaware to analyse behavior in independent settings. They used three error resolution strategies: clarification-audio, clarification-visual, clarification-audio-visual. The Wizard-of-Oz experiment was conducted using two screens. One showed participant view of the system and the other showed the

participant sitting in a sofa. The Wizard was located in an adjacent room where the study was conducted.

Interaction for the application was design based on the low-fidelity prototypes created. The buttons on each page were listed and the links and the functions were defined. Any interactive functions that were difficult to implement were also made note of to either use during the development of personalised interface or future work.

4.8 Interface properties

The properties on the various pages of the interface need defining before they are designed or developed in the relevant applications.

Usability is defined in ISO 9241 part 11 (a standard giving guidance on usability on requirements for office work with visual display terminals) as the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use. Although usability can survive without multimedia, it is not possible for multimedia to achieve its full potential without the efficient use of usability [12]. Hvannberg and Law came up with defect classifications to improve product quality and motivation process improvement [13]. Classification of usability problems (CUP) scheme comprises ten attributes, which are: identifier, description, defect removal activity, trigger, impact, expected phase, failure qualifier, cause, actual phase and error prevention.

Hummels and his team stated that resonance of users while interacting with a product should be a key issue in the development of human computer interaction [14]. This unites research areas and capitalises their strength creating pleasant and attractive products. They went on to develop a product called Image and Sound Handling (ISH), which was an interactive multimedia system which had the following characteristics: Gatherish, Compose, Sitish, Smallish, Stirish, Jitterish, Acoutish, Rhythmish and Mixish. This experiment proved that resonance was an individual experience which enhanced HCI principles. Baranauskas and Scolari state that HCI is a discipline, which involves the knowledge of both computer science and human nature; Semiotics combines both these domains [15]. Semiotics is not just the signs of signs but a theory about sense

production and interpretation. Semiotic analysis includes software interfaces, video game interfaces and narrative of interactions.

Tano and colleagues state that it has gradually become obvious that the latest information systems have a serious drawback in creating promotive work [16]. They went on to explain such drawbacks using an experiment called "Godzilla" that supports creative design. They discovered problems such as ignorance of suitable cognitive mode and ignorance of suitable means of information representation. Sturm and Mueller developed a product for a foreign market with a target group with different cultural background and its challenges [17]. They addressed the differences provided by cross-cultural research in anthropology, linguistics and psychology. HCI and Usability issues will be of prime consideration in interface development.

Hongpaisanwiwat and Lewis state that animated characters captured user's attention, engaged them in tasks and entertained them [18]. Such techniques contribute to effective multimedia presentations. Quality of voice has also been an important ingredient of comprehension. This team investigated the influence of introversion – extroversion on effectiveness of presentations and found that animated characters did not increase learning but proved to be useful to retain the learner's attention and maintain engagement with learning material. Oh and Stuerzlinger presented a new conceptual 3D design system that manipulated composite objects having considered alternative manipulation techniques to separate complex paths of an existing models [19]. They presented two new techniques that supported separation of objects into complex parts: intelligent separation (user clicks with a left mouse button on a block and drags it away to separate a part) and separation with anchoring (here the user must first specify part that remain in place). This was a novel approach for easy modification of a conceptual 3D design. Both 2D and 3D animation are to be a major part of the interface properties.

X-AiD is an intelligent shell used to design and operate user interfaces. It offers an integrated and extendable structure of icons, windows, menus, etc. Creation of these interfaces are carried out by following a set of syntax, semantics, heuristic and display rules [20]. Although this is an application used generally for

interface creation this cannot be customised to create personalised interfaces as needed for this research.

Text is to be used to indicate the active sound in the interface. The buttons for the sounds also contain text in Arial font. The active sound would be displayed in around font size 48 and the other button text would be font size 18. The buttons would have a reasonable space in-between. The navigation indicating replay and exit are also to be around font size 18. In total there would be 12 sound buttons, the active sound image and the two navigation buttons.

Various shapes were considered for the shapes of buttons and the interface objects [21]. Rectangle was chosen having researched into Gestalt Laws and psychology of visual perception [22]. Previous investigation 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 [23], 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 orange turning into red on rollover. The active sound would also be dark yellow. The colours for the individual 2D animation would be restricted to a maximum of three colours. The 3D animation colours are natural and do not change.

Still images are not to be used in this interface. There will be two moving images, one demonstrating the sound made and the other indicating an example of the sound made. This will attract the attention of the child and illustrate movement of the mouth to demonstrate the sound that is made.

There will be two types of audio used for the interface. An introductory sentence with instruction to click on buttons and a thank you on exit would be used. For each demonstration of sound, the sound produced would be played.

4.9 High Fidelity Prototyping

Based on the evaluation of the low fidelity prototypes, interaction design and interface property specification and the high fidelity prototypes were made.

Modeling, texture and lighting, animation and camera and rendering of clips using the audio files of the twelve chosen sounds, based on the organs used in formation for each of these sound, the relevant parts of the mouth were designed and developed for both the boy and a girl. Files of the individual animations for each sound in both Macromedia Flash and Maxon's Cinema 4D software were created and optimised for Flash and published. The skeleton of the application was designed in Flash with three Scenes. Buttons were prepared and space was allocated for each of its contents. Finally, the animation files were inserted along with the relevant Actionscript.

The Flash animation was exported as a shockwave format and the C4D file was exported to Combustion and with the audio files exported as MPEG4 format. This interface was evaluated by the focus group in an iterative process for development.

4.10 Evaluation

Each version of the interface was evaluated by members of the focus group using ethnomethodology, phenomenology and action research before being experimented with participants. Suggestions were made as to the development of the main page of the interface and different versions of the interface and interaction were developed iteratively (figure 6). Textual analysis was also used to redefine requirements for the interface.

Hornbæk and Frøkjær stated that understanding human thinking is crucial in the design and evaluation of human computer interaction inspired by introspective psychology [24]. They presented five metaphors of human thinking namely, metaphor of habit formation, metaphor of the stream of thought, metaphor of the dynamic thinking, metaphor of the incompleteness of utterance, metaphor of knowing. These five metaphors were considered when the focus group evaluated the interfaces when they were developed.

Children with normal mental/cognitive abilities are often introduced to cross-sections of anatomy at the age of 6. It is not possible to say if disabled children aged 7-8 or younger would understand the concept or may find it scary [25]. It was suggested that the sounds 'n', 's', 'm' are

not released as early stages of development emphasise on the firm contact to produce the sound and not on the release. The difference between voiced and unvoiced sounds also had to be distinct.



Figure 6 - iterative versions

Five amendments were needed for the model:

- Shape of uvula needs to be more defined – should look more like a droplet
- Uvula must have more mobility – close and open for the different sounds (always closed except for nasals)
- Teeth – cross section would be realistic without the teeth
- Tongue – should be clearer in demonstrating the tip, blade, centre, back of tongue
- It can be hard to see the position of the tongue for certain sound as the teeth cover them.

Changes were needed for the animations:

- S - tongue should be clearly placed behind the teeth
- M – do not release sound keep lips closed
- N – don't release the sound, keep tongue positioned behind teeth
- Ck – back of tongue should be up against uvula and push the air off – nothing should happen to the throat
- t/d – may have to get rid of the lower teeth for demo and the difference between the unvoiced 't' and the voiced 'd' should be clear.
- H – should be unvoiced with only air passing through
- P – shuts for 'p' and opens after - contact is more important than the release
- R – roll tongue inside mouth but don't roar. The tip of the tongue is lower than usual positions
- a/e/i – check the different positions or lifting of the tongue.

'Hit' areas of the buttons need to be defined. As it is, each text has to be selected. Children may be able to visualise images better if they are cartoon based rather than looking realistic and scary. This may look more medical than a game. Minimise, Maximise and Restore facility similar to windows would be useful or some sort of off/on functionality. This was made note of for phase two.

Intro to the main page could have a sound indicating exp. 'click on the letter to hear the sound' and similarly the exit page could mean having a voice over. Eg. 'Goodbye'. A reward system would be useful even if it means the therapist has to play the 'good' or 'try again' rewards. 3D window may need to have two

perspectives showing the shaping of the mouth (front and side) in order to emphasis the shaping of the mouth from the front as well.

The voice needs to be more welcoming preferably with an upper inflection that welcomes children to come and come and join. The sounds used are very tired and sound difficult. Font should be changed into 'Sasoon primary' which is the standard font used for speech/sound training and it should be lower case

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

Modeling, texture and lighting, animation and camera and rendering of clips using the audio files of the twelve chosen sounds, based on the organs used in formation for each of these sound, the relevant parts of the mouth were designed and developed for both the boy and a girl. Files of the individual animations for each sound in both Flash and C4D were created and optimised for Flash and published. The skeleton of the application was designed in Flash with three Scenes. Buttons were prepared and space was allocated for each of its contents. Finally, the animation files were inserted along with the relevant ActionScript.

The Flash animation was exported as a shockwave file and the C4D file was exported to Combustion and with the audio files exported as MPEG4.

In phase two, 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 (figure 7).

The designing of the preferences option in the interface was evaluated by the focus group and changes were made iteratively. To cater to the

visually impaired and different types of disability with different levels of distractively options in changing the window size and turning in completely off were given. The instructors voice was also an option to choose from a male of female voice.



Figure 7 – Final version of preferences

Barendregt and his team identified difficulties encountered by evaluators when conducting formative evaluation in usability [26]. They conducted formative evaluation of both usability and fun in computer games for children. Their first study suggested negative effects of tasks for fun assessments. The second study showed difficulties in determining set of heuristics. Although this interface had a fun element there was an option to minimise or exit the window which was distracting the user.

An interactive application for children with special needs assessing usability and accessibility issues when using websites were designed by Perdrix and his team [27]. The results showed that children aged 7-10 carried out their tasks individually but did not lose any details once the functions were explained. Children older than 10 did not read the text and did not pay attention to what could be done but located the most attractive aspects such as games or chat. Children between 3-6 years old who have not yet begun to read or just beginning to do so needed visual interfaces and also the text had to be brief. Although this interface had various windows to cater for different age groups, an option to minimise or exit the window which was distracting the user eg. An older child may not need the 2D animation for examples.

A help page was included to indicate on rollover what each option meant (figure 8).

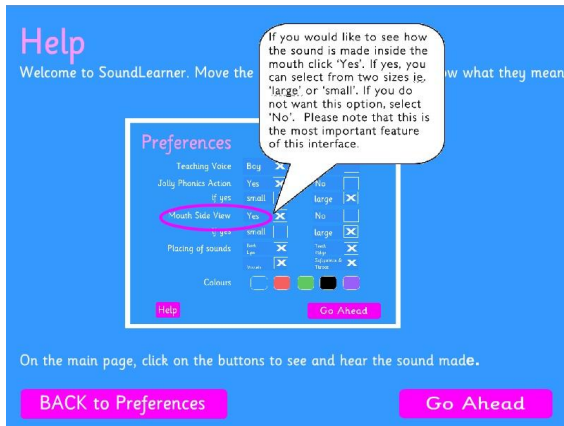


Figure 8 – Final version of help page

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

5 Experiments

Phase one of the research lasted 6 weeks and produced the final group interface. Phase two of the research lasted 8 weeks and produced the final personalised interface. Experiment 01 was carried out following the low fidelity prototype development during evaluation, experiment 02 was carried out during phase 01 and experiment 03 was carried out during phase 02. Statistical analysis was carried out and conclusions were reached.

Interfaces using adoptive learning algorithms can be used for designing interfaces. This system enables personalised profiles of interfaces based on algorithms to reflect the users disability and their individual interactions with the system. Algorithms are to be written during the evaluation of the interfaces. Garneau and Holynski stated as computer interfaces or graphical displays become more sophisticated and easy to generate, we must remove the influences of programmers and add the individuality of the users [28]. This requires applications to generate algorithms to create personalised graphical displays. Measuring an

individual's personal preferences is by no means an easy task. There are infinite numbers of factors such as background, gender, expertise, etc. The system must be able to determine user preferences quickly and effectively and use decisive rules to generate new graphic displays. Creating individual preferences is the key requirement when creating personalised interfaces. This research intends to create personalised interface quickly and effectively by doing extensive experiments with a focus group of able participants before using the interface with the disabled participants.

Bohórquez and colleagues carried out a survey of AAC tools in Columbia [29]. They tried to answer the question 'How to enable the handicapped in a reasonable environment so that they can achieve higher levels of human development'. This team aimed to generate supporting devices in order to make the users autonomous and interactive actors in different given environments in a natural way. They used a research methodology for device development and the setting of AAC as well as interdisciplinary and inter-institutional processes. They concluded that generalisation and homogenisation of the population is not permissible given that the individual must be considered. Every child or adult have their own level of problem due to their environment. These differences have to be considered in order to design high or low-level technology devices. Scherer and his team attempted to validate AT baselines and outcomes to quantify the best match of consumer and AT [30]. This study was set in a prospective multi-cohort environment. The study was carried out with around 150 carers in 25 states with one consumer receiving a new AT. The results showed there were no significant differences due to gender, physical loyalty or age within this sample of working adult consumers. This team also used an evaluation form to make assessments every six months.

Participants between ages 5 years to 36 years, 4 females and 6 males, 7 able and 3 disabled evaluated the interfaces (Table 2).

QuickTime™ and a
 TIFF (Uncompressed) decompressor
 are needed to see this picture.

Table 2 – Participants of experiments

5.1 Experiment 1

This phase of the study investigated;

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. The results obtained are shown in 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

Participant	Number of Attempts per Sound									
	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 3 – Attempts taken per sound with no interface

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;

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

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. The number of attempts taken per sound were recorded (Table 4).

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 4 - Attempts taken per sound with group interface

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 (Tables 6,7).

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.

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 6 - No interface

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

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

Table 7 - Group 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.

can be used as a starting point when evaluating the interface with the participants.

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

5.3 Experiment 3

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

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 preferences window was used to control the dimensions of the window depending on the priority of need. The number of attempts taken per sound were recorded (Table 5).

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	2	pending
a	2	3	3	3	2	4	2	2	2	2	
ck	4	3	3	3	2	4	3	2	2	2	
d	2	2	4	4	2	3	4	2	2	2	
e	2	4	3	3	2	3	4	4	2	2	
h	2	3	3	3	2	5	3	2	2	2	
I	2	2	3	3	2	2	3	2	2	2	
m	2	2	3	3	2	2	3	2	2	2	
n	2	2	3	2	2	6	3	4	2	2	
p	3	3	4	3	3	3	4	4	3	3	
r	2	2	3	3	2	7	5	3	2	2	
t	3	3	4	4	3	4	4	4	2	2	

Table 5 - Attempts taken per sound with personalised interface

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

Table 8 - Personalised interface

T-test was conducted on the two sets of data (Tables 7, 8). 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 the tables indicated that the performance of the personalised interface was much better than the group interface.

The graph in figure 9 shows the number of attempts made to pronounce each sound using the various interfaces. The graph 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 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.

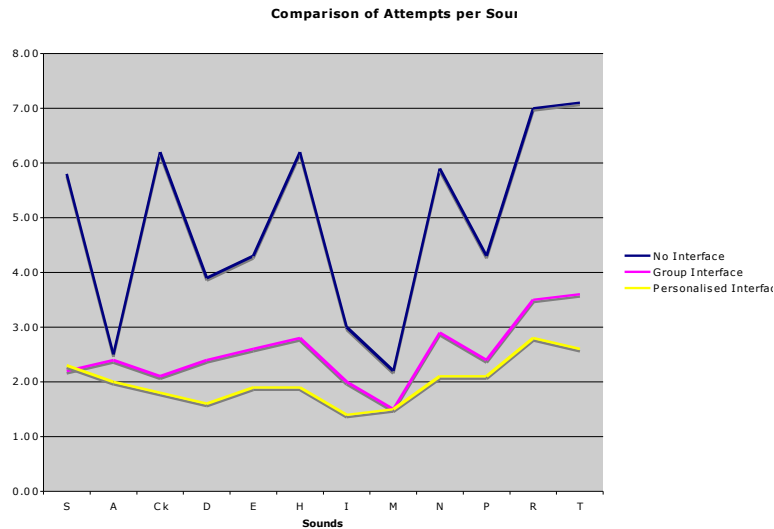


Figure 9 – Comparison of average attempts

6 Conclusions

Having established the need for a multimedia interface to enhance pronunciation, the research problem on the type of interface needed had to be addressed. The research problem to be addressed 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?”

Two interfaces were developed to solve the research problem. 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 results clearly show that having any interface was better than no interface and the personalised interface offered a better success rate than the group interface as indicated by t-test and comparison of the number of attempts to pronounce a sound. Thus 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 main stream schools. Although there have been many attempts to create an interface to enhance pronunciation none has been completed successfully. This fact was emphasised by the speech experienced 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.

The interfaces could be integrated into mobile devices [31]. It could also be used as a medical tool for diagnosis and further treatment [32]. Further research is expected to be done to move this milestone even further.

Acknowledgement

Pauline Philips of Speech and Language Therapy, members of the focus group, participants.

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] George, J., Gnanayutham, P., (2007), Research Methodology for Phonological Development, HCI International 2007, Lecture Notes in Computer Science Series, Posters, Springer, July 2007, Beijing, 907 – 911.
- [3] Vaknin, S., (2005), In Defense of Psychoanalysis, *The Fundamentals of Psychological Theories*, Narcissus Publications.
- [4] George, J., (2007), Theories for Educational Multimedia Application Development: A Case Study, HCI International 2007, Lecture Notes in Computer Science Series, Posters, Springer, July 2007, Beijing, 1211 – 1215.
- [5] Gimson, A.C., (1962), *An Introduction to the Pronunciation of English*, Edward Arnold Publications.
- [6] JollyLearning (2006),
<http://www.jollylearning.co.uk/jp.htm>, Accessed 27th July 2006
- [7] Ager, S (2006),
<http://www.omniglot.com/writing/ipa.htm>, Omniglot, Accessed 29th August 2006.
- [8] Passy, P., (1888), *International Phonetic Alphabet*, International Phonetic Association, France.
- [9] Draheim, D., Weber, G., (2003),
Storyboarding form-based interfaces, Interact 2003, IOS Press, Zurich, Switzerland (343-350).
- [10] Holzinger, A., Ebner, M., (2003), *Interaction and usability of simulations and animations: A case study of the Flash technology*, Interact 2003 (777-780), Zurich, Switzerland, Published by IOS Press.
- [11] Berglund, A., Qvarfordt, P., (2003), *Error resolution strategies for interactive television speech interfaces*, Interact 2003, IOS Press, Zurich, Switzerland (105-112).
- [12] George, J.C., (2006), *Usability issues in Multimedia Interface Design*, ATINER, Athens June, 2006.
- [13] Hvannberg, E. E., Law, L, (2003),
Classification of Usability Problem (CUP) scheme, Interact 2003, IOS Press, Zurich, Switzerland (655-662).
- [14] Hummels, C., Ross, P., Overbeeke, K. C. J., (2003), *In search of resonant human computer interaction: Building and testing aesthetic installation*, Interact 2003, IOS Press, Zurich, Switzerland (399-406).
- [15] Baranauskas, C., Scolari, C. A., (2003),
Semiotics of Interaction: a special interest group

proposal, Interact 2003, IOS Press, Zurich, Switzerland (1095).

[16] Tano, S., Kodera, T., Nakashima, T., Kawano, I., Nakanishi, K., Hamagishi, G., Inoue, M., Watanabe, A., Okamoto, T., Kawagoe, K., Kaneko, K., Hotta, T., Tatsuoka, M., (2003), *Godzilla: Seamless 2D and 3D Sketch Environment for Reflective and Creative Design Work*, Interact 2003, IOS Press, Zurich, Switzerland (311-318).

[17] Sturm, C., Mueller, C. H., (2003), *Putting theory into practice: How to apply cross-cultural differences to user interface design?* Interact 2003, IOS Press, Zurich, Switzerland (1051-1052).

[18] Hongpaisanwiwat, C., Lewis, M., (2003), *Attentional effect of animated character*, Interact 2003 (423-430), Zurich, Switzerland, Published by IOS Press.

[19] Oh, J., Stuerzlinger, W., (2003), *Intelligent manipulation techniques for conceptual 3D design*, Interact 2003 (319-326), Zurich, Switzerland, Published by IOS Press.

[20] Hein, H., Kellermann, G. M., Thomas, C. G., (1987), *X-AiD: A shell for building highly interactive and adoptive user interfaces*, IUCA 1987, August 1987, Milan, Italy (97-99).

[21] 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

[22] Pickford, R. W., (1972), *Psychology and Visual Aesthetics*, Hutchinson Educational, London, 17 – 30.

[23] 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.

[24] Hornbæk, K., Frøkjær, E., (2003), *Metaphors of human thinking: a new tool in user interface design and evaluation*, Interact 2003, IOS Press, Zurich, Switzerland (781-784).

[25] Philips, P., (2006), *Speech and Language Therapy*, Interviewed on 7th June 2006 by author.

[26] Barendregt, W., Bekker, M. M., Speerstra, M., (2003), *Empirical evaluation of usability and fun in computer games for children*, Interact 2003, IOS Press, Zurich, Switzerland (343-350).

[27] Perdrix, F., Granollers, T., Lorés, J., Zubillaga, A., (2003), *Usability evaluation applied to a children's website*, Interact 2003, IOS Press, Zurich, Switzerland (852-855).

[28] Garneau, R.W., Holynski, M., (1989), *The interactive adoptability of graphic displays*, Sallvendy, G., Smith, M., (Eds.), *Designing and using human-computer-interfaces and knowledge based systems*, Elsevier Science Publishers, Amsterdam, 1989, (321-327).

[29] Bohorquez, J.C., Gonzalez, M., Varela, A., Macias, H., Rivera, J., Rodriguez, N., Gonzalez, M., Toro, I., (2004), *Supporting Technology for Augmentative and Alternative Communication*, CWUAAT, 2004, Cambridge, UK.

[30] Stone, D, Jarrett, C, Woodroffe, M, Minocha, S (2005), *User Interface Design and Evaluation*, The Open University, UK: Elsevier, Morgan Kaufmann Publishers.

[31] Reeves N, Jefferies L, Cunningham S-J. And Harris C. (2007), *A multimedia PDA/PC speech and language therapy tool for patients with Aphasia*, International Journal of Electronic Healthcare. Vol. 3(1) p. 135 - 149.

[32] Glykas M. and Chytas P. (2004) *Technology assisted speech and language therapy*, International Journal of Medical Informatics. 73 p.529-541.