SpeechKit: A Multimedia Speech Tool

David Calder
Curtin University of Technology
Bentley, Perth,
West Australia
Tel. 61-8-9266 2875
david.calder@cbs.curtin.edu.au

ABSTRACT

SpeechKit is a multimedia system for use by speech therapists to assist in the rehabilitation of motor impaired patients. These are often stroke victims who have to relearn the ability to communicate effectively. The process is laborious and usually conducted in a one-to-one situation where interaction between therapist and client depends on visual and audio cues. Instead of the usual paper-based prompts used by the therapist, this multimedia solution offers screen prompts together with high quality digitised speech. The result is a reduced work load for the therapist. The system has undergone beta tests in Western Australia. The well established and robust Asymetrix Toolbook was used for development of the speech therapy tool, the current version in mid 2008 being version 9.1.

Keywords

Multimedia, Speech Impaired, Speech Therapy, Cardiovascular Accident, Computer-Based Therapy.

1. INTRODUCTION

Strokes (cardiovascular accidents) are common and extremely traumatic for any society as well as on a personal level. Those that survive this trauma usually have need for rehabilitation. Motor impairment of the speech organ is a common associative occurrence. There is a worldwide shortage of speech training services to accommodate the clients who require speech therapy [1]. We have produced a Multimedia system which replaces the chart and paper-based cues found in most speech therapy units. The aim is to assist speech therapists by decreasing the time they have to spend in direct contact with patients. More importantly, there are also a number of benefits to clients and these will be discussed later. We worked closely with speech professionals from Royal Perth Hospital in order to tailor the interface exactly to their requirements. We were guided by practicing speech therapists in ensuring existing and well established interactive therapy strategies for certain classes of dysarthria, were followed as closely as possible. The first version of the system has undergone beta testing in Western Australia.

2. SPEECH LOSS DUE TO STROKE

A Stroke or Cerebrovascular accident (C.V.A) occurs when the oxygen supply to the brain is blocked or when an artery in the brain is ruptured. The former is the result of a clot travelling to the brain and occluding blood flow in a cerebral artery [2]. Once the oxygen supply has ceased, hydrogen ions propagate within the brain cells and damage the cells. Within four to eight minutes, the cells begin to die, and of course this destruction is irreversible. In the case of a reduced blood flow, cells may stop functioning but later recover when normal perfusion is restored [3].

Cardiovascular accidents are the third largest killer in Australia and in many other countries of the western world. The highest killer is heart disease followed by cancer [4]. Fifteen percent of people over the age of forty living in the western world die of cardiovascular accidents. A further fifteen percent are also stroke victims who survive and require some form of rehabilitation [5]. Of this group five percent make a full recovery whilst ten percent suffer some permanent disability [6].

As medical technologies have improved and the relative death rate has fallen, so the need to offer suitable rehabilitation services to the victims who have survived this trauma has increased. Patients are left with multiple disabilities and these often include loss of speech. The anterior portion of the brain is associated with speech production and damage to this area can result in motor speech disorders [7]. The severity of the motor speech disorder may vary from person to person but even where spontaneous recovery is evident, some speech rehabilitation will probably be carried out. In the case of more severe speech impairment, long term rehabilitation may result [8]. In most instances, the road to recovery is long and difficult, placing stress on both client and therapist. Any means that may assist this load for both the client and the therapist should be investigated.

3. PAPER-BASED PATIENT CUES

During a regular session, the client is presented with a number of cues which include symbols representing phonemes/vowels, drawings indicating air flow through the mouth and nose and the many combinations of lip, teeth and tongue positions which are essential in general articulate speech. Cards such as the one shown in Figure 1 were identified as part of a major problem by speech professionals during the initial data analysis of the old system at Royal Perth Hospital, Shenton Park. The therapist is under pressure to find the correct flash card or phoneme chart as she sits with the client. Together with these paper-based cues she uses her own speech and face to offer sound and visual cues to the

client. In order to produce the consonant M for example, the client may be asked to place the lips together and allow the air to flow through the nasal cavity. The therapist will give an example by producing the sound herself and simultaneously accentuate the positioning of the lips.

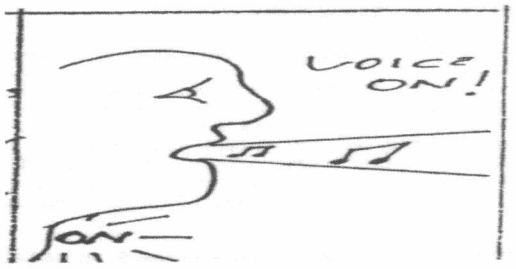

Figure 1. A basic hand-drawn cue card.

The client is then asked to repeat this step. Consonant and consonant/vowel combinations such as this need constant repetition by the client and consequently the therapist before any improvement is expected. This face-to-face contact can be very tiring for both the client and the therapist. As therapy is done on a one to one basis, duration of contact per individual per week may be limited to only a few hours. Clients are therefore being deprived of the continuous training which would benefit their rehabilitation.

4. MULTIMEDIA THERAPY SOLUTION

The concept of a system to assist in speech therapy resulted from earlier work with speech therapists in the United Kingdom in developing a speech aid prototype [8]. The development of SpeechKit depended on a requirement to address some of the most significant problems associated with conventional therapy methods. The Department of Speech Pathology at the Royal Perth (Rehabilitation) Hospital, Australia, was involved in the development of the system and provided the old model of operations for analysis from which the computer-based system was developed. Assistant Professor G. Mann, now in the Department of Communication Disorders at the University of Florida and widely recognised as a key authority, worked closely with us over several years in developing the original version of the multimedia tool [10].

SpeechKit aims to relieve constant therapist/patient supervision, particularly where time consuming repetitive tasks are involved. Therapists can use their time more effectively in planning new goals whilst the computer provides visual and sound cues to the client. Therapists no longer have to organise cue cards or sort through hundreds of icons and drawings. These were seldom in colour whereas the computer-based system augments these traditional methods by using colour and animation. The latter was something that could not be achieved on loose pieces of paper or cardboard! Consequently the therapy process can run more smoothly and effectively as all cues are presented on the screen and/or produced by the high quality stored speech system.

There are a number of other benefits which particularly relate to patients and have been highlighted during beta testing. It has been found that the system could be used at home as well as in the conventional environment of the speech therapy unit. Where a client is left with the computer and removed from the clinical surroundings, the stress of embarrassment is removed. This is particularly evident when an older client has to "relearn how to speak" in the presence and under the direction of a young speech therapist. The stress placed on the client during these encounters should be seriously considered.

Another bonus for the system is that it is based on a standard IBM compatible PC and could even be run from a portable notebook mounted to a wheelchair. Most households now have several computers with a more than adequate built-in sound systems [11]. This means that in most instances, only the software need be installed in the home to allow therapy to continue. Other members of the family could be involved in the rehabilitation and the clinical sessions at the hospital could be used for the monitoring of progress and not be associated with intrusion into the privacy of the patient [12]. Communication within the family is an important part of rehabilitation and SpeechKit could help in promoting this. The high cost of therapy is relieved but not replaced by this augmentative system [13].

Since the SpeechKit system saves clinician time, it may also help relieve the shortage of speech therapy services. If the amount of time spent with each client is reduced, the clinician could then take on a greater case load.

The nature of a computer-based system is that, "it doesn't have to see another patient in an hours' time". Therefore clients may train at their own pace under no stress. Sound and visual cues may be repeated over and over again without the pressure of supervision. Of course not all patients would be suitable candidates for this level of freedom but initial tests have indicated great success for certain motor impaired victims.

5. THE TOOLBOOK SOLUTION

Asymetrix Multimedia Toolbook has been used to develop this system. It operates in a high level, Windows-based object-oriented programming environment. Toolbook uses a book metaphor for multi screen production. Each screen is described as a page and all pages in a production are called a book. Making use of these properties pages were constructed to emulate the icons and visual cues used in conventional therapy (see Figure 2).

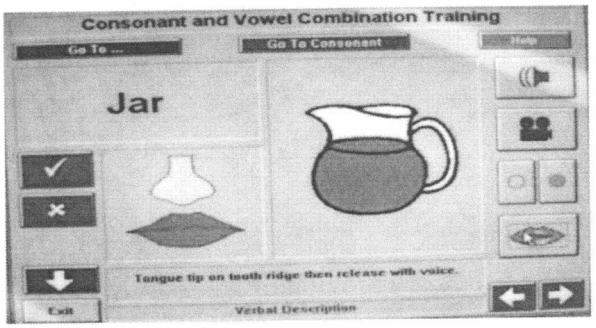

Figure 2. A SpeechKit Interface Display

Animated visual representations of the vocal system were added to the standard set of cues. High quality digitised speech was recorded from a practicing speech therapist so that an exact representation of a set of speech sounds could be achieved. As the normal therapy session involves a progressive set of prompts to initiate speech from the client, SpeechKit was set up with a progressive set of cue buttons which gradually give more and more assistance to the client in achieving a particular goal. For example, this might be help in pronouncing the consonant F. A set of six cue buttons mounted vertically on the right hand side of the screen allow for this positive reinforcement. These cues may take visual or spoken form. Another reason for a progressive interactive strategy is the problem of complexity which has plagued so many previous interface designs for the disabled [14].

The sound cues may be both instructive and exemplary. One cue button will produce an instruction such as, "place the lower lip against your upper front teeth and blow" whilst another produces the target as required, the actual sound "fff ....". Another button links the consonant to a word so as to place the target sound in context. Yet another offers the word in the context of a sentence.

All cue buttons and command buttons have explanatory logos which appear at the bottom of the screen when the cursor moves over their area. There are over 300 digitised sound recordings associated with SpeechKit.

The system is normally controlled by a mouse and test models incorporate this means of user control. Other remote switch devices are currently being investigated. This will make the system more versatile and therefore suitable for quadriplegics [15]. A system of this type should be flexible enough to adapt to suit the needs of each individual client and varying physical disabilities [16].

Research by the authors has established that phoneme synthesis is not suitable for the speech requirements of this system. A robotic sounding voice may be acceptable for certain games programs and simple communication tools but can never be adequate for the variety of accents, languages and key sounds that may be required by therapists. The Institution of Electrical and Electronics Engineers [17], published a recommended practice for speech quality measurements in 1969, but this has never been fully adopted as the basis for standardised testing [18]. Edwards showed that no existing rule-based synthesiser came close to passing the test of being indistinguishable from natural speech.

To date, the system has been tested on a small pilot sample of patients in Australia. Initial tests at the Royal Perth Hospital have indicated that the response to the system for middle age stroke victims was very positive. Other age groups including children are included in forward planning for the later version. The interface has proved easy to understand and selected stroke patients have managed to control the system themselves, i.e. choosing and activating cue buttons. Future tests will be coordinated to fine tune the system before full trials take place.

6. FUTURE DIRECTIONS

Feedback from the first pilot trials highlighted the need for fast adaptability in design refinement which could be partly addressed by the digital ecosystem paradigm [19]. Issues such as universal access are also relevant [20].

Assistive device design challenges are compounded if the user has multiple disabilities [21]. The design team, who may be focussing exclusively on one area of disability, does not always consider these multi-facetted demands. This problem is not confined to speech tools. Multiple disabilities can be physical, cognitive or both. Issues of complexity with respect to individual requirements must be seen within the context of a wider ecology of the particular user, with that person clearly at the centre, contributing to a team solution. An established and highly successful ecological model for the above can be found in the design of individualized education programmes for the disabled. This has been refined over twenty years into a highly recommended model and is now regarded as 'best practice' [22]. The ecological approach has not as yet permeated all areas of disability support. However, the power of the digital ecosystem framework is now accepted within many other disciplines [23].

The logic behind collaborative workflows is to produce a sequence of activities that not only produce a meaningful result, but also to facilitate small groups working together to achieve common goals. The actual physical distance and associated limitations between these entities then becomes less important as web based tools are used to link enterprises and their common aspirations [24]. The entities themselves may typically be widely dispersed cottage industries (such as those associated with the design and manufacture of assistive devices and therapy tools for the disabled user). The key point is they have a common interest [25].

It is believed that multimedia solutions such as SpeechKit need to be embedded in a co-operative digital ecosystem that is user

centred, benefiting not only the patient, but also the therapist and support teams and the cottage industry that designs the software.

7. CONCLUSIONS

SpeechKit is a multimedia-based training tool for use by speech therapists and their clients. Based on findings from Beta tests carried out in Western Australia, A more advanced system is currently under development. This new version is specifically aimed at people with motor speech disorders but interest has been shown by groups who work with young cerebral palsy victims, and that has been taken into account. It is believed that it may, in time, be developed into a wider range of products which may incorporate related therapy requirements. The immediate aim is to market the newer version of the system and allow further research and development to progress accordingly.

8. REFERENCES

[1] Editor, 2008, May 22. ASU Designs Program To Address Shortage of Speech Therapists, Phoenix Arizona News. DOI= http//evliving.com

[2] Hewer, R.L.& Wade D.T. 1986. Stroke - A Practical Guide Towards Recovery. Methuen Australia : Australia.

[3] Kaplan N.M. & Stamler M.D. 1983. Prevention of Coronary Heart Disease. W.B. Saunders and Co.

[4] Gross, P.F. 1991 The Total Economic Costs of Stroke in Australia. Technology Assessment Corporation Pty Ltd. NSW, Australia.

[5] Hird, K.M. 1987 Communicative Competence, Self-care and Mobility in Left and Right Hemisphere Adult Stroke Patients. Curtin University of Technology, Western Australia.

[6] Dyken, M.L., Woff P.A., Barnett J.H. 1984. Risk Factors in Stroke: A Statement for Physicians by the subcommittee on Risk Factors and Strokes. Stroke (6): 1105-1111.

[7] Darley, F.l., Aronson A.E, Brown J.R 1975. Motor Speech Disorders. W.B. Saunders Co. : London.

[8] Gloag, D. 1985 Rehabilitation After Stroke: What is the potential? British Medical Journal 290, 699-701.

[9] Calder, D.J. , Lister P.F., Mack Watson, J.D 1987 Large Vocabulary Speech Aid Incorporating Softkey Switching. Third Symposium International De Ingeniera Biomedica. Madrid. 755 - 781.

[10] Mann. G. 2002. The Mann Assessment of Swallowing Ability. Cengage Delmar Learning, New York.

[11] Edwards A., Pitt I. 2002. Design of Speech-Based Devices. Practitioner's Series. Springer Professional Computing, London.

[12] Alkalay, L & Asserman M.W. 1983. Cerebrovascular Accidents: Care and Rehabilitation. The Practioner 227. 469 - 473.

[13] Hodge S. 2007. "Why is the potential of augmentative and alternative communication not being realized? Exploring the experiences of people who use communication aids". Disability and Society, 22, 466

[14] Seidle, F. 1977. Is Home Care Less Expensive? Health Social Work 2 (5) - 10.

[15]] Vanderheiden, G.C and Lloyd, L.L. 1986. Communication Systems and Their Components. Augmentative and Alternative Communication: American Speech and Language-Hearing Association. Rockville (Maryland) 49 , 162.

[16] Reichle, J.J. York,J., Sigafoos J 1991 Implementing Augmentative and Alternative Communication. Paul Brookes Publishing Co. London 239 - 256.

[17] IEEE 1969. IEEE Recommended Practice For Speech Quality Measurements. IEEE Transactions on Audio and Electronics, AU,17 (3), 225 - 246.

[18] Edwards,A.D. 1991. Speech Synthesis: Technology for Disabled People. Paul Chapman Publishing Ltd: London 12 - 40.

[19] E. Chang, M. West. 2006. "Digital Ecosystems and Comparison to Collaboration Environment". WSEAS Transactions on Environment and development 2, 1396-1404

[20] Zimmermann, G,, Vanderheiden, G., Gilman, A. 2002. Prototype Implementations for a Universal Remote Console Specification. CHI 2002 Conference on Human Factors in Computing Systems, Minneapolis, MN, pp 510 – 511.

[21] J. Jacko, H. Bautsch, 2000. "A Conceptual Framework For The Integration of Clinical Diagnosis, Functional Capabilities and Classes of Technology", University of Wisconsin Lab for Human-Computer Interaction. Technical Report series no. 1131, 1 – 59

[22] B. Rainforth, J. York, C. Macdonald, 1993. Collaborative Teams for Students With Severe Disabilities, Baltimore: Paul Brookes, 71- 83.

[23] L. Pudhota, Chang E, 2004. "Modelling the Dynamic Relationships between Workflow Components" ICEISI Porto, Portugal,.

[24] M. UlieruR. Brennan Scott, 2000. "The Holonic enterprise: a model for Internet-enabled Global Manufacturing Supply Chain and workflow Management", Canada.

[25] E. Chang, T. Dillon and F Hussain, 2006. " Trust and Reputation for service-oriented Environments: Technology for building Business intelligence and Consumer Confidence", John Wiley and Sons, West Sussex, England,