SPECS – an embedded platform, speech-driven environmental control system evaluated in a virtuous circle framework.

Heidi Christensen[1,2], *Siddharth Sehgal*[2], *Peter O'Neill*[2], *Zoë Clarke*[3], *Simon Judge*[3], *Stuart Cunningham*[2], *Phil Green*[1],*Mark Hawley*[2,3]

[1]Department of Computer Science, University of Sheffield, UK
[2]School of Health and Related Research, University of Sheffield, UK
[3]Department of Medical Physics and Clinical Engineering, Barnsley District General Hospital, UK

h.christensen@dcs.shef.ac.uk,s.sehgal@sheffield.ac.uk,mark.hawley@sheffield.ac.uk

Abstract

The aim of the SPECS project is to implement a speech-driven environmental control (EC) system for elderly and disabled people. This type of assistive technology (AT) enables users to control electronic and electrical devices in their homes in situations when using conventional means of access, such as remote controls or out-of-reach light switches, is not possible or desirable.

Some commercial voice-operated EC systems exists alongside more traditional switch-operated systems. However, they are only supporting very limited command-style speech, and in general, commercial systems employing speech technology (for both the non-AT and AT market) are all designed for users with typical speech[1]. Unfortunately, users in need of AT often have physical disabilities associated with motor control and such conditions (e.g. cerebral palsy) will also likely to affect the musculature surrounding the articulatory system resulting in slurred and less clear speech; known as *dysarthric* speech.

Research into the use of natural language enabled interfaces have demonstrated their power, and new projects like the recently funded UK programme grant: Natural Speech Technology[1] plan to apply state-of-the-art speech research to the assistive technology domain for both typical and dysarthric speakers.

Dysarthric speakers' recognition performance, when using commercial ASR systems, can be very poor because of the increased variability of the speech coupled with the often very limited amount of training material available. The SPECS project focuses on building high-performance ASR systems for dysarthric users, as well as devising a framework for ensuring an EC system which is tailored to the needs of the individual[2, 3].

In SPECS, we build on experience from a previous AT related project in Sheffield, Stardust. One of the major outcomes of that project was the demonstration that the conventional enrolment phase, where the users are asked to record a number of examples of each command word, can be boosted by adding a *user training* phase[4, 5]. In this phase the user practises how to pronounce the command words in a game-like environment; the outcome is two-fold: i) the user will over time learn which pronunciations are preferred by the underlying acoustic models, and ii) as all the speech commands are recorded, a substantial amount of additional audio is collected.

Another main focus of the SPECS work has been to devise a plan for how best to incrementally improve the whole system including the ASR element. This has seen the idea of the *virtuous circle*, which is an established framework in the therapeutic community, coupled with higher performance acoustic models and menu configurations. For each user, our *virtuous circle* contains an enrolment stage, a system trial stage and an evaluation stage. During the enrolment stage, the user's needs are assessed and speech training material is recorded both through dedicated recording sessions and via user training sessions. A full system is then configured on the SPECS device with appropriately trained acoustic models and grammars, and the device is deployed in the user's home for a trial.

Central to the *virtuous circle* framework is the idea that the performance of the system can be analysed at suitable milestones during the evaluations. After a period of trialling the system in the home, the performance is evaluated. The trialling stage is expected to increase in time for each iteration starting from as low as a week. A number of evaluation methods have been used ranging from traditional ASR word recognition rates, to analysing "in-the-wild" system performance from log files and the accompanying audio, and to running offline system simulations.

The SPECS device itself is purpose build for the project by Toby Churchill Ltd[2]. It is based on a balloon 3 board and runs emDebian. It has a 320 (width) x 240 (height) pixel LCD screen and comes with build in infrared support via an integrated GEWA PROG III micro chip. The restrictions imposed on the software are mainly in terms of fixed-point computation, memory usage and real-time issues. As a result, all of the SPECS software including acoustic model training and decoding, recording and user training module, and GUI has been written specifically for the project.

In the current configuration, the ASR system is *push-to-speak* meaning that the user has to push a switch every time he or she wishes to issue a command. The SPECS ASR system uses standard speaker dependant, whole word, left-to-right HMMs trained with MFCC features. The decoder uses context-dependant grammars that are switched at runtime.

So far four participants have been enrolled in the SPECS evaluations. They are split between two typical speakers and two dysarthric speakers. The user evaluations will take the form of case studies, and they are comprised of three stages with the fourth and final stage in progress: a) enrolment, b) user training, c) System A evaluations, d) System B evaluations, where System A and System B denotes the first two iterations of systems in our virtuous circle.

[1] http://www.natural-speech-technology.org

[2] http://www.toby-churchill.com

Overall, the results are very encouraging with increasing recognition performances through all the completed stages of the evaluations for all users. For the three users who have currently completed the first evaluation, the obtained recognition results are 92.70% (dysarthric speech user), 99.70% (typical speech) and 97.60% (typical speech) respectively, averaging 96.7%.

1. References

[1] K. Rosen and S. Yampolsky, "Automatic speech recognition and a review of its functioning with dysarthric speech," *Augmentative & Alternative Communication*, vol. 16, no. 1, p. 4860, Jan 2000.

[2] M. Hawley, S. Cunningham, S. Judge, B. Kolluru, and Z. Robertson, "Using qualitative research methods to inform user centred design of an innovative assistive technology device," in *4th Cambridge Workshop on Universal Access and Assistive Technology*, Cambridge,UK, April 2008.

[3] S. Judge, Z. Robertson, M. Hawley, and P. Enderby, "Speech-driven environmental control systems - a qualitative analysis of users' perceptions," *Disability and Rehabilitation: Assistive Technology*, vol. 4, no. 3, pp. 151–157, 2009.

[4] M. Parker, S. Cunningham, P. Enderby, M. Hawley, and P. Green, "Automatic speech recognition and training for severely dysarthric users of assistive technology the STARDUST project," *Clinical Linguistics and Phonetic*, vol. 20, no. 2-3, pp. 149–156, 2006.

[5] M. Hawley, P. Enderby, P. Green, S. Cunningham, S. Brownsell, J. Carmichael, M. Parker, A. Hatzis, P. ONeill, , and R. Palmer, "A speech-controlled environmental control system for people with severe dysarthria," *Medical Engineering & Physics*, vol. 29, no. 5, pp. 586–93, 2007.