Conference paper

Hawley, Mark, Cunningham, Stuart, Cardinaux, Fabien, Coy, Andre, O'Neill, Peter, Seghal, Siddharth and Enderby, Pam (2007) *Challenges in developing a voice input voice output communication aid for people with severe dysarthria.* In: AAATE 2007 : 9th European Conference for the Advancement of Assistive Technology in Europe, Oct 2007, San Sebastian, Spain.

Challenges in developing a Voice Input Voice Output Communication Aid for People with Severe Dysarthria

Mark HAWLEY[a], Stuart CUNNINGHAM[d], Fabien CARDINAUX[a], Andre COY[a], Peter O'NEILL[a], Siddharth SEGHAL[c], Pam ENDERBY[b]
[a]*Department of Medical Physics and Clinical Engineering, Barnsley Hospital, UK*
[b]*School of Health and Related Research, University of Sheffield, UK*
[c]*Department of Computer Sciences, University of Sheffield, UK*
[d]*Department of Human Communication Sciences, University of Sheffield, UK*

Abstract. This paper describes the development of a voice-input voice-output communication aid (VIVOCA) for people with disordered or unintelligible speech, initially concentrating on people with severe dysarthria. The VIVOCA is intended to recognize and interpret an individual's disordered speech and speak out an equivalent message in clear synthesized speech. User consultation suggests that such a device would be acceptable and would be useful in communication situations where speed and intelligibility are crucial. Speech recognition techniques build on previously successful development of speech-based home control interfaces, and various methods for speech 'translation' have been evaluated.

Keywords. Speech recognition, voice output communication aid, dysarthria

Introduction

Dysarthria is the most common acquired speech disorder affecting 170 per 100,000 population [1]. It is present in approx 33% of all people with traumatic brain injury, 8-20% of people with cerebral palsy and increases in incidence with the progression of neurological diseases such as MND and MS. Severe dysarthric speech is often unintelligible to unfamiliar communication partners. As a result, people with dysarthria often use voice output communication aids (VOCAs), but these devices have limitations. Despite much research and development, VOCAs are often slow and tiring to use. Because of their reliance on a switch or keyboard and screen for user interfacing, they do not facilitate natural communication with eye contact between conversation partners. Indeed, many people with disordered speech prefer to speak, even if it is effortful and fails, as it is quicker and more immediately responsive than any other communication method.

A communication method is required that retains, as far as is possible, the speed, naturalness and responsiveness of speech communication but, crucially, adds to intelligibility. This paper describes the development of a voice-input voice-output communication aid (VIVOCA). The VIVOCA is intended to recognize and interpret an individual’s disordered speech and speak out an equivalent message in clear synthesized speech. Figure 1 shows a schematic of the device and its major components.

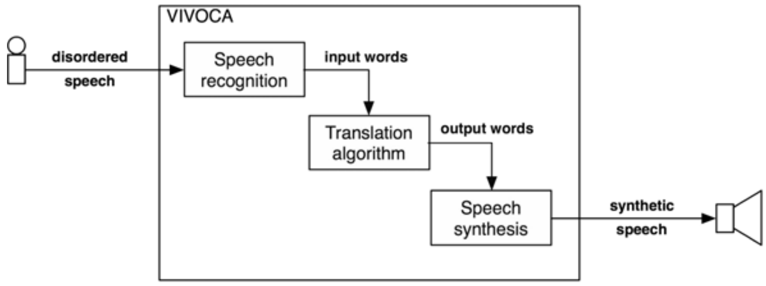

Figure 1. System diagram of the voice-input voice-output communication aid (VIVOCA).

1. User views and requirements

We consulted VOCA users, including potential VIVOCA users for their views on the acceptability and potential utility of a VIVOCA. Eight of the 12 users consulted reported that they would be prepared to use a VIVOCA. They perceived potential benefits to be increased speed of communication and reduced keyboard use. The users conveyed the idea that the VIVOCA might increase the ability to communicate, increasing self-expression and independence. Specific situations that more than one user would like to use a VIVOCA for include meeting new people, talking on the telephone and shopping; all situations where speed and intelligibility are crucial.

Figure 2.One of the developers demonstrating the prototype device.

Consultation with users indicated the need for a portable, if possible body-worn, communication device, but with a sufficiently large screen so that user and communication partner could view it. Consequently a pocket computer (PDA) was chosen as the hardware platform. Figure 2 shows a prototype of the device, with input via a blue tooth headset, in use by one of its developers.

2. Speech recognition

The speech recognition techniques used in this project build on previously successful development of speech-based home control interfaces [2,3], where accurate recognition of highly variable, severely dysarthric speech in noisy environments was shown to be feasible. However, whilst home control tasks can be carried out with a relatively small number of control inputs (small input vocabulary), speech communication requires a larger input vocabulary. Small vocabulary recognition accuracy reported in the literature varies from 22% to 81%, under test conditions, for severe dysarthria [4-8]. We have previously found that, although recognition accuracies above 80% are consistently attainable for small vocabularies of dysarthric speech [3], recognition accuracy decreases rapidly with increasing input vocabulary. The development of VIVOCA has required new techniques for optimising recognition accuracy. Figure 3 shows a comparison of the performance of a conventional modelling technique with and without the pre-processing of speech prior to model building. Pre-processing dysarthric speech combined with model optimisation results in a much less rapid deterioration in recognition accuracy with increasing input vocabulary. Using these techniques, we have achieved acceptable recognition accuracy at vocabulary sizes approaching 100 words, and it is expected that this can be increased still further.

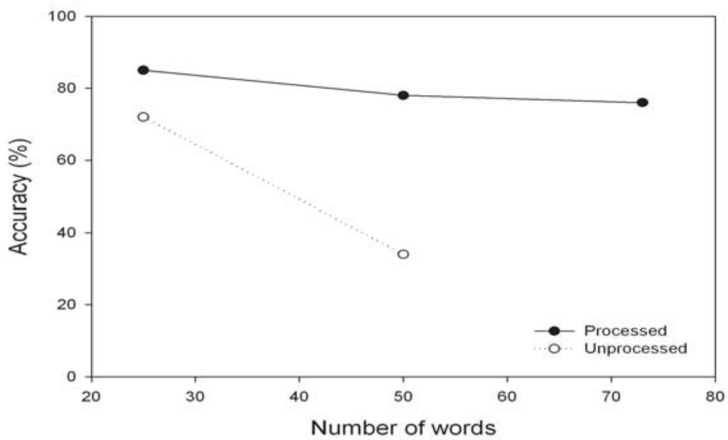

Figure 3. Accuracy of speech recognition for different sized input vocabularies for a single speaker with severe dysarthria.

3. Translation

Once the input speech of the user is recognised, the challenge is to 'translate' this input into the required output. Despite optimisation of recognition, available input vocabularies are restricted and vary from individual to individual. A number of methods for translating input speech into output speech have therefore been explored and modelled, including direct word to word translation, spelling (using the NATO alphabet), AAC techniques using symbolic grids, in addition to novel methods of building phrases from word combinations. Selecting the method for the device requires optimisation with respect to the available input vocabulary, the desired output vocabulary and the rate of communication. Other considerations include cognitive and perceptual load on the user. Figure 4 shows a graph of rate of communication against recognition accuracy for a number of different potential ‘translation’ schema, derived from mathematical models. Rate of communication is greater for techniques that require less input to produce a given output, as expected, but falls off rapidly with decreasing input recognition accuracy, due to the time spent by the user in correcting errors. This modelling has allowed us to explore the trade-offs between input vocabulary and recognition accuracy with respect to communication rate. Potential users have been involved in working through these trade-offs in order to design a suitable translation method and user interface.

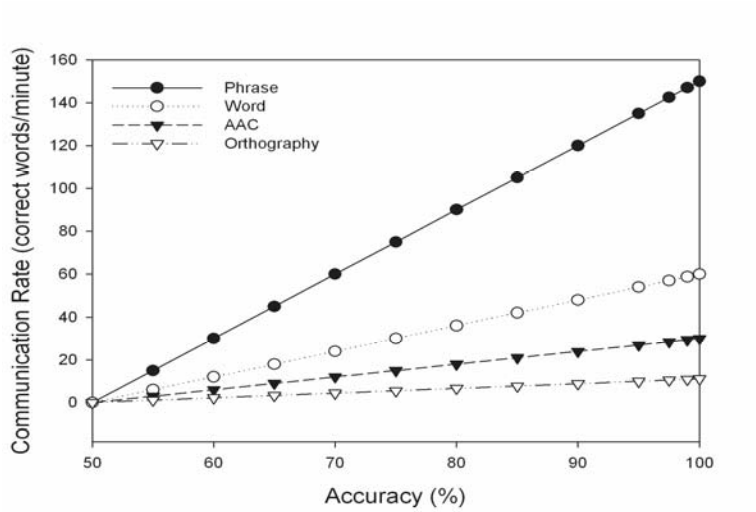

Figure 4. The effect of recognition accuracy on communication rate for different ‘translation’ methods. Phrase is the phrase building technique used in VIVOCA, Word is direct word-to-word translation, AAC is a symbol combination technique used in a popular VOCA, and orthography is spelling of each word.

4. Implementation on a mobile device

According to the user requirements, the communication device should be portable and attractive. Moreover the users showed a wide interest in having both visual and audio feedback and being able to correct their last entry if they make a mistake or if the communication device misrecognises their input. All these requirements raise many technical challenges.

We chose to use a PDA as the hardware platform since it is a low cost portable device with an appropriate screen, and it allows the software to be implemented using standard programming tools. Implementation of a high complexity algorithm such as automatic speech recognition (ASR) in a low computation, low memory device is a very challenging task. In particular, a PDA does not include hardware for floating-point computation. Most PDAs will emulate floating-point calculations in software, however for ASR this is not a viable solution given the significant time overhead. We addressed this problem by implementing the ASR algorithms using fixed-point calculations. As a result, a short sequence of input words can now be recognised in less that 1 second.

The implemented algorithm also allows the user to get visual and audio feedback (in an ear-piece for example) from the last pronounced word, and to delete it if necessary. Only when a full sequence of words is finished, does the device synthesize the translated sentence and speak it out. In addition, the user can read on the device screen the list of all possible words that he/she can speak.

5. Conclusions

A speech recognition engine for dysarthric speech has been produced to run on a PDA and is delivering acceptable performance for even very severely dysarthric speakers. Techniques for translating recognised word strings into synthesised speech output have also been created. A VIVOCA prototype, designed according to user requirements, has now been produced and is being trialled by users in order to optimise the design prior to a larger trial.

Acknowledgements

This research was supported by the UK Department of Health New and Emerging Applications of Technology programme.

References

[1] Enderby, P and Emerson, L (1995) Does Speech and Language Therapy Work? Singular Publications p.84.

[2] Hawley MS, Enderby P, Green P, Brownsell S, Hatzis A, Parker M, Carmichael J, Cunningham S, O'Neill P, Palmer R, (2003) "STARDUST; Speech Training And Recognition for Dysarthric Users of Assistive Technology." In Assistive Technology – Shaping the Future, eds: GM Craddock et al, IOS Press, pp 959-64.

[3] Hawley, M.S., Enderby, P., Green, P., Cunningham, S.P., Brownsell, S. Carmichael, J., Parker, M., Hatzis, A., O'Neill, P., and Palmer, R., A speech-controlled environmental control system for people with severe dysarthria, Medical Engineering and Physics 29, 5 (2007) 586–593

[4] Hawley MS, Speech Recognition as an Input to Electronic Assistive Technology, British Journal of Occupational Therapy, 2002, 65(1), 15-20

[5] Rosengren E, Raghavendra P, Hunnicut S How does automatic speech recognition handle severely dysarthric speech? Proc 2nd TIDE Congress, Paris, April 1995

[6] Thomas-Stonell N, Kotler A-L, Leeper HA, Doyle C Computerized speech recognition: influence of intelligibility and perceptual consistency on recognition accuracy, Journal of Augmentative and Alternative Communication, 1998, 14, 51-55

[7] Ferrier LJ. Shane HC. Ballard HF. Carpenter T. Benoit A. Dysarthric Speakers' Intelligibility and Speech Characteristics in Relation to Computer Speech Recognition. Journal of Augmentative and Alternative Communication, 1995, 11, 165-174

[8] Kotler, A., Thomas-Stonell, N Effects of speech training on the accuracy of speech recognition for an individual with a speech impairment, Journal of Augmentative and Alternative Communication, 1997, 12: 71-80.