IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 1, Issue 6, Dec-Jan, 2014
ISSN: 2320 - 8791
www.ijreat.org

A New Form Of Augmentative And Alternative Device For People With Severe Speech Impairment

Preethi.S[1], Arun prasath.J[2] (AP)

[1]ECE Department, Anna University, Prathyusha Institute of technology and Management, Trivallur, Tamilnadu,India

[2]ECE Department, Anna University, Prathyusha Institute of technology and Management, Trivallur, Tamilnadu,India

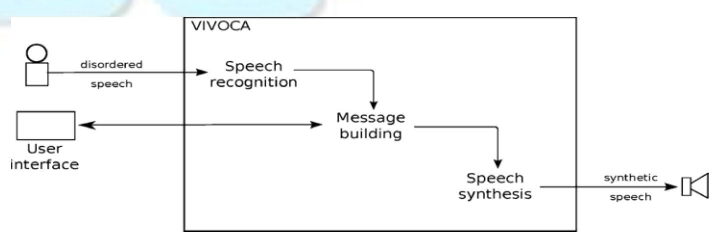
ABSTRACT: The voice-input voice-output communication aid (VIVOCA) is a new form of augmentative and alternative communication (AAC) device for people with severe speech impairment. The VIVOCA recognizes the disordered speech of the user and it builds messages, which are converted into synthetic speech. System development is a user-centered design, the development methods are identified and refined. Speaker-dependent automatic speech recognizers are used for building small vocabulary with reduced amount of training data is used. Experimentally Mean accuracy of 96% is achieved for highly disordered speech and good recognition performance is also obtained even when recognition complexity is increased. The selected message-building technique includes various factors like speed of message construction and range of available message outputs. The VIVOCA was evaluated with the individuals having moderate to severe dysarthria and the results showed that they can make use of the device to produce intelligible speech output from disordered speech input.

Keywords **-** *Augmentative and alternative communication, automatic speech recognition, dysarthria, voice output communication aid.*

I INTRODUCTION

Speech impairment is often associated with severe physical disabilities as a result of progressive neurological conditions such as motor neurone disease, congenital conditions such as cerebral palsy, or acquired neurological conditions as a result of stroke or traumatic brain injury. Current technological tools for communication, voice-output communication aids (VOCAs), generally rely on a switch or keyboard for input. Consequently, they can be difficult to use and tiring for many users, and they do not readily facilitate natural communication as they are relatively slow and disrupt eye contact report that users need a device which is physically easy to operate in a wide range of positions and environments. Many people with VOCAs often prefer to speak rather than use the aid, even if their speech is largely unintelligible, as it is a more natural form of communication. In addition, found that listeners rated users of a communication aid as more socially competent if they had a more rapid rate of delivery. It is there- fore desirable that a new communication aid retain, as far as possible, the speed and, ideally, the naturalness of spoken communication.

For speech to be seen as a viable means of interacting with assistive technology for people with dysarthria, it must be shown to function well in the home environment, where conditions of ambient noise and microphone positioning can be far from optimal for speech recognition. Moreover, speech control must demonstrate advantages over interfaces currently used for environmental control for people with severe physical disabilities, the most common of which is switch scanning



In this work, we applied statistical ASR techniques, based on hidden Markov models (HMMs), to the speech of severely dysarthric speakers to produce speaker dependent recognition models, and developed a novel methodology for recognizer-building. This approach relied on a user-training phase in which the user practiced speaking to the recognizer, whilst receiving consistent visual feedback based on the similarity between their current attempt and the distribution of their previous attempts. This enabled the user to become more efficient at producing the target utterances, by reducing variation in their vocalizations, while at the same time facilitating the collection of additional speech examples that were then used to train the final recognizer. These enhancements resulted in speech recognition being a viable means of controlling assistive technology for small input vocabularies, even for people with severe speech disorders.

II SYSTEM DESCRIPTION

The development made use of a user-centered design. An initial detailed user requirements study considered the views of both potential VIVOCA users and of speech and language therapists/pathologists who provide voice- output communication aids. A wide range of user requirements were elicited and the VIVOCA was implemented to meet these requirements where feasible. The development process was iterative and the implementation was gradually refined by testing developments.

III WORKING OPERATION

The VIVOCA works as per the following description. The step by step operation process are given as follows.

A. SPEECH RECOGNITION

In prevailing methods, automatic speech recognition (ASR) is based on statistical models (usually HMMs) of speech units. For a large vocabulary system, the speech units will be at the level of individual speech sounds, phones. The resulting speaker-independent recognizer can be adapted for an individual speaker, given a small amount of enrolment speech data from that speaker. However, this ASR technique is unsuitable for speakers with severe speech disorders because the amount of material available for training is severely limited (as speaking often requires great effort), the material is highly variable. Instead, we have introduced a new methodology for building small vocabulary, speaker-dependent personal recognizers with reduced amounts of training data.

The recordings consist of isolated productions of each of the words that are required for the recognizer's input vocabulary. These examples are used to train the initial whole word models. In this study we used HMMs with 11 states, with astraight-through arrangement. The acoustic vectors were 12Mel-frequency cepstral coefficients (MFCCs) derived from a 26-channel filter bank with a 25 ms analysis window and 10ms frame-rate. Energy normalization and cepstral mean normalization were also applied to the input features. This is a conventional ASR front-end. The models were trained using the HMM toolkit with the Baum-Welch algorithm.

This approach is straightforward for a typical speaker, but it is more problematic for the intended users of a VIVOCA due to their speech impairment. This means there is a scarcity of training data and the consequences of this are exacerbated by the variability in the productions of speakers with dysarthria. This application prompts the user repeatedly to speak each of the words in the initial recognition vocabulary. Each utterance is recorded, but crucially the user is given feedback on "closeness of fit" of each attempt to their own recognition model.

At the conclusion of this user-training step, the recognition models can be re-estimated using both the initial training examples, and subsequent examples collected with the user training application, producing a recognizer which is more accurate and robust to variations in the user's speech. The process can, of course, be repeated. We have previously found that recognition accuracies above 80% for isolated words, and above 70% for commands (short strings of words) are consistently attainable for small vocabularies of severely dysarthric speech. Whilst home control tasks can be carried out with a relatively small number of control inputs (and small input vocabulary of around 10–15 words), supporting speech communication requires more flexibility in its output and is therefore likely to require a larger input vocabulary. For speaker-dependent recognition it is known that word recognition accuracy falls with increasing vocabulary size, and this reduction is likely to be exacerbated when speech input is highly variable, as is the case with dysarthric speech.

B. MESSAGE BUILDING

The message building module constructs messages, which the user wishes to communicate, from the recognized input words. The simplest, and in many ways the ideal, form of message building, given that we are recognizing word units, would be to recognize

ISSN: 2320 - 8791
www.ijreat.org

each word individually and speak out the same word in a clearer (synthesized) voice. As part of the user-centered design and development process, design meetings were held between the research team and potential users at which different message-building methods were considered, with knowledge of the modelled communication rates. Users prioritized

methods which tended to have high communication rate. Some users, however, also regarded a large output vocabulary as being vital regardless of its effect on communication rate.

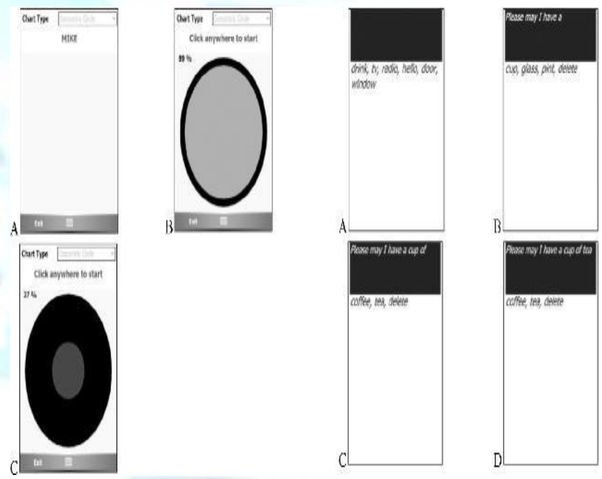
Phrase building is used to generate frequently used phrases requiring rapid generation, such as answering the phone, conversational fillers or communicating immediate needs/problems. For example, inputting the sequence of words "want" "drink" "water" could generate the phrase "Can I have a drink of water please." Using this approach in a structured way greatly reduces the recognition perplexity. Spelling may be used for the remainder of less-frequently used words allowing unlimited output vocabulary where greater precision and conversational range are required, though at the cost of greater perplexity and much lower communication rate.

C. SPEECH SYNTHESIS

The system software was de- signed to work with both prerecorded output (in the form of waveform files) and to interface with a speech synthesizer. As with the process of speech recognition, speech synthesis is a computationally demanding process. The PDA takes voice input from the user via a microphone, or the internal microphone of the PDA. The PDA's internal speaker was found to produce speech at too low a volume for practical use in any but quiet ambient conditions. Therefore, a separate amplifier and speaker were used for the spoken output. The central processing units (CPUs) in PDAs do not have support for rapid numerical computation, and have no dedicated hardware for floating-point calculations.

III HARDWARE AND SOFTWARE IMPLEMENTATION

In order to meet users' requirements, the hardware upon which VIVOCA was implemented needed to be small and light and have a suitable visual interface.



A VOICE-INPUT VOICE-OUTPUT COMMUNICATION AID FOR PEOPLE WITH SEVERE SPEECH IMPAIRMENT

IV CONCLUSION

This paper has described the development of portable, voice output communication aid controllable by automatic speech recognition. The device can be configured to enable the user to create either simple or complex messages using a combination of a relatively small set of input "words." Evaluation with a group of potential users showed that they can make use of the device to produce intelligible speech output. The evaluation also, however, highlighted several issues which limit the performance and usability of the device, confirming that further work is required before it becomes an acceptable tool for people with moderate to severe dysarthria. Overcoming these limitations will be the focus of our future research.

V REFERENCES

[1] D. Beukelman and P. Mirenda, *Augmentative and Alternative Communication*, 3rd ed. Baltimore, MD: Paul H. Brookes, 2005.
[2] Blaney B, Wilson J. Acoustic variability in dysarthria and computer speech recognition. Clin Linguist Phonet 2000;14(4):307–27.
[3] P. Enderby and L. Emerson, *Does Speech and Language Therapy Work?*. London, U.K.: Singular, 1995.

[4] Thomas-Stonell N, Kotler A-L, Leeper HA, Doyle C. Computerized speech recognition: influence of intelligibility and perceptual consistency on recognition accuracy. J Augment Altern Communication 1998;14:51–5.
[5]C. L. Kleinke, "Gaze and eye contact: A research review,"*Psychol. Bull.*, vol. 100, no. 1, pp. 78–100, 1986.
[6]B. O'Keefe, N. Kozak, and R. Schuller, "Research priorities in augmentative and alternative communication as identified by people who use AAC and their facilitators," *Augmentative Alternative Commun.*, vol. 23, no. 1, pp. 89–96, 2007.
[7] Hawley MS. Speech recognition as an input to electronic assistive technology. Brit J Occup Ther 2002; 65(1):1520.
[8]J. Murphy, "I prefer contact this close: Perceptions of AAC by people with motor neurone disease and their communication partners," *Augmentative Alternative Commun.*, vol. 20, pp. 259–271, 2004.
[9] Rosengren E, Raghavendra P, Hunnicut S. How does automatic speech recognition handle severely dysarthric speech? Proceedings of the second TIDE congress. 1995.
[10]U. Sandler and Y. Sonnenblick, "A system for recognition and transla- tion of the speech of handicapped individuals," in *9th Mediterranean Electrotech. Conf. (MELECON'98)*, 1998, pp. 16–19.
[11]B. Wisenburn and D. J. Higginbotham, "An AAC application using speaking partner speech recognition to automatically produce contex- tually relevant utterances: Objective results," *Augmentative Alternative Commun.*, vol. 24, no. 2, pp. 100–109, 2008.
[12]B. Wisenburn and D. J. Higginbotham, "Participant evaluations of rate and communication efficacy of an AAC application using natural lan- guage processing," *Augmentative Alternative Commun.*, vol. 25, no. 2, pp. 78–89, 2009.
[13]M. S. Hawley *et al.*, "A speech-controlled environmental control system for people with severe dysarthria," *Med. Eng. Phys.*, vol. 29, no. 5, pp. 586–593, 2007.
[14]L. J. Ferrier, H. C. Shane, H. F. Ballard, T. Carpenter, and A. Benoit, "Dysarthric speakers' intelligibility and speech characteristics in relation to computer speech recognition," *Augmentative Alternative Commun.*, vol. 11, no. 3, pp. 165–175, 1995.
[15]N. Thomas-Stonell, A. L. Kotler, H. A. Leeper, and P. C. Doyle, "Com- puterized speech recognition: Influence of intelligibility and percep- tual consistency on recognition accuracy," *Augmentative Alternative Commun.*, vol. 14, no. 1, pp. 51–56, 1998.