A breath controlled AAC system

A Breath Controlled AAC System

D. Kerr, K. Bouazza-Marouf, A. Gaur*, A. Sutton and R. Green
Wolfson School, Loughborough University
*University Hospitals of Leicester NHS Trust

An estimated 2 million people in the USA and from 0.3% - 1.0% of the total world population of school-age children have a need for an Augmentative and Alternative Communication (AAC) intervention of some kind (Beukelman & Mirenda 2013).

There is a vast range of people with Complex Communication Needs (CCN), from those with chronic or degenerative conditions, those requiring immediate post-surgical intervention, or those suffering traumatic brain injury. Such conditions can also include Dysarthria (disorder of motor speech control), Apraxia (disorder of coordination of motor movements for speech), or Aphasia (language disorder as a result of brain injury or cerebral vascular incident). With such a wide range of conditions, and of the varying physical and cognitive abilities associated with them, it is inevitable that AAC systems have to be carefully tailored to individual needs.

Michael Williams (Williams et al. 2008) defines “Five Principles for the next 25 years of AAC“. Amongst these, they emphasise the need for systems to be versatile enough to support communication modes that enable participation in all aspects of 21st century life. Further, they specify that users should be fully involved in the selection, training and adjustment of the device or devices they are using, thereby taking ownership of their communication system. However, modern devices still have a long way to go to satisfy some of these principles.

So-called “high-tech” AAC methods use direct selection or scanning interfaces to provide speech output. Storage and retrieval methods are often slow (anything from 3 to 12 words per minute). Narrative conversation is thus very difficult (“normal” conversational speech is 270 wpm). Built-in vocabularies may lack the flexibility required for all but the simplest exchange of ideas. A comprehensive study (Hodge 2007) of AAC users found that one of the major obstacles to successful uptake was the physical interface, where a user’s movement impairments often meant frustratingly slow and inefficient response times. This was compounded by the need for assistance from carers or relatives to pre-programme the machine with chosen words or phrases.

Activation input must be carefully tailored to an individual user (e.g. eye gaze, sip/puff etc.). Sip/puff systems already exist where breath is used to activate a switch (both simple and advanced). A breath-controlled encoding device (Dilbagi 2014) was unveiled at a recent Google Science Fair, and a breath-controlled switch has been introduced to regulate “textual flow” in the Dasher communication system developed by Cambridge University (Tom H. Shorrock, David J. C. MacKay 2004). An example of breath signals we collected using a digital microphone is shown in Figure 1.

However, the digital codes produced from signals like this, in order to guide a screen cursor to make fixed selections, are slow, laborious and limited in direct information content. Humans are not really designed to select, think or speak in binary codes!

Analogue methods of encoding are more “natural” and potentially can contain more information. For example, small variations in amplitude can be used to represent nuance, emphasis or emotion. Phase changes might represent grammar variations, such as tense, plurals, or queries. This compares to the semantic compaction methods used in Minspeak and Blissymbols, where such “prosodic features” of language can currently be represented within an AAC system. Greater communication speeds should thus be achievable, thanks to the more rich information content possible. Some analogue signals with differing phases can be seen in the example of Figure 3. One could for example think of these as past and present tense versions of the same word or phrase.

For this project, we started out to design a simple, needs-driven communication device for patients in a hospital ICU on ventilator support. Later, we sought to extend this more generally to anyone with a severe speech impairment, which may be combined with partial or complete loss of motor functionality. Our starting assumption was that all such people were able to breathe spontaneously; hence we could use breath as an activation method. However, potential users:

- must be able to control breathing rate spontaneously
- must have sufficient cognitive ability to understand how to use system
- in later stages of degenerative diseases may not be suited to the device

A schematic of our device is shown in Figure 2, and some typical breathing pressure/time signals or “patterns” obtained from it are shown in Figure 3. Breathing variations within the user’s mask are detected by a pressure sensor, and then captured as patterns by a computer program and analysed for characteristic features, such as variations in frequency, intensity and phase, as depicted in Figure 3.

The computer-based recognition system then “learns” to understand these patterns through example and repetition. The data once recognised can be turned into speech using either voice synthesis or pre-recorded playback. Users have to “teach” the system to recognise the breath patterns that they wish to associate with words, phrases etc. There are thus two modes of use.

- Learning mode – user “teaches” the system new words or phrases by providing examples. Vocabulary is built up as required over time.
- Speaking mode – system recognises patterns breathed by user and turns them into synthesised speech, or plays back pre-recorded voice.

We tested our prototype system on 7 healthy people aged 20-22 years, with no speech impairment. Each provided 10 repetitions of 3 unique breath patterns of their choice. After a short period, each was asked to reproduce their 3 unique patterns five times each for the system to interpret. The mean reliability of system interpretation was found to be 90% (values ranged between 73% and 100%)

As the development of our system proceeds, the standard approach of a control interface to some bounded selection set is what we want to get away from. We want to impose as few limitations as possible on what the control interface can achieve in terms of communicating wants, ideas, feelings etc. in the form of narrative communication. We intend to incorporate Natural Language Processing and Generation into our system (to provide contextual data, disambiguate meaning and complete broken sentences, for example).

In the future, we will look at Artificial Intelligence (AI) applications that would allow better interfaces between users and their machines (reducing critical reliance on family and/or care workers to train and maintain the system). The objective would be an intelligent AAC system that works with its user to grow and develop a means of narrative-enabled communication suited to the individual’s needs and capabilities, without the requirement for constant intervention by others.

Two very interesting developments in AI point towards what might be possible in the future. The Talking Heads Experiment (Steels 2015) was the first large-scale trial in which artificial agents created for the first time a new shared vocabulary by playing language games about real world scenes in front of them. If such machine to machine communication is possible, then human to machine language might also be designed to evolve in a natural way, complete with vocabulary, syntax, morphology, etc. and set within an appropriate contextual basis.

Intelligent tutoring systems, such as AutoTutor (Graesser et al. 2005), developed at the Institute of Intelligent Systems, University of Memphis, are designed to simulate a human tutor’s behaviour and guidance, through a dialogue with the student. The student’s understanding and depth of learning are enhanced and guided by the tutoring system, using so-called “conversational agents” founded on constructivist learning theories. This type of system might be designed to guide the new AAC user through the learning process and the development of the most suitable language structure to suit their needs. The system could thus grow to suit the capabilities of the speaker, and evolve over time as conditions change.

We think that this is an exciting time for AAC development, where a new generation of devices using recent advances in engineering and cognitive science may one day be possible, that are truly adaptive to an individual’s needs, and that may go even further in satisfying Michael Williams’ five excellent principles.

Acknowledgements

We would like to acknowledge the help and support of the University Hospitals of Leicester NHS Trust, who have provided funds for equipment for the project.

References

Beukelman, D.R. & Mirenda, P., 2013. *AAC: Supporting Children and Adults with Complex Communication Needs* 4th ed., Paul Brookes.

Dilbagi, A., 2014. Talk breathes new life into the alternative communication device market. Available at: http://newatlas.com/talk-morse-code-speech/33042/.

Graesser, A. et al., 2005. AutoTutor: an intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*, 48(4), pp.612–618.

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 & Society*, 22(5), pp.457–471. Available at: http://www.tandfonline.com/doi/abs/10.1080/09687590701427552 [Accessed August 19, 2016].

Steels, L., 2015. *The Talking Heads Experiment*, Freie Universitat Berlin. Available at: http://langsci-press.org/catalog/book/49.

Tom H. Shorrock, David J. C. MacKay, C.J.B., 2004. Efficient Communication by Breathing. In N. L. Joab Winkler, Mahesan Niranjan, ed. *Deterministic and Statistical Methods in Machine Learning*. Springer, pp. 88–97.

Williams, M., Krezman, C. & McNaughton, D., 2008. “Reach for the Stars”: Five Principles for the next 25 years of AAC. *Augmentative and Alternative Communication*, 24(3), pp.194–206. Available at: http://search.ebscohost.com/login.aspx?direct=true&db=c8h&AN=2010055637&site=ehost-live.

Figures

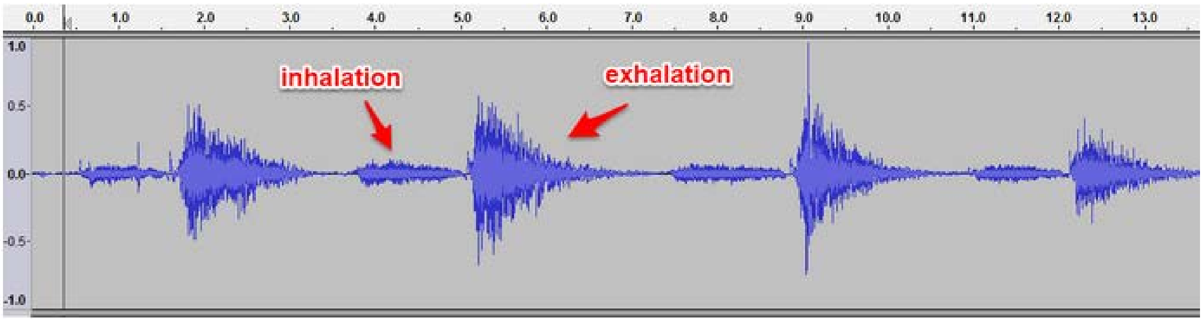

Figure 1 – sound signals recorded by a cardioid digital microphone placed adjacent to the mouth during normal breathing.

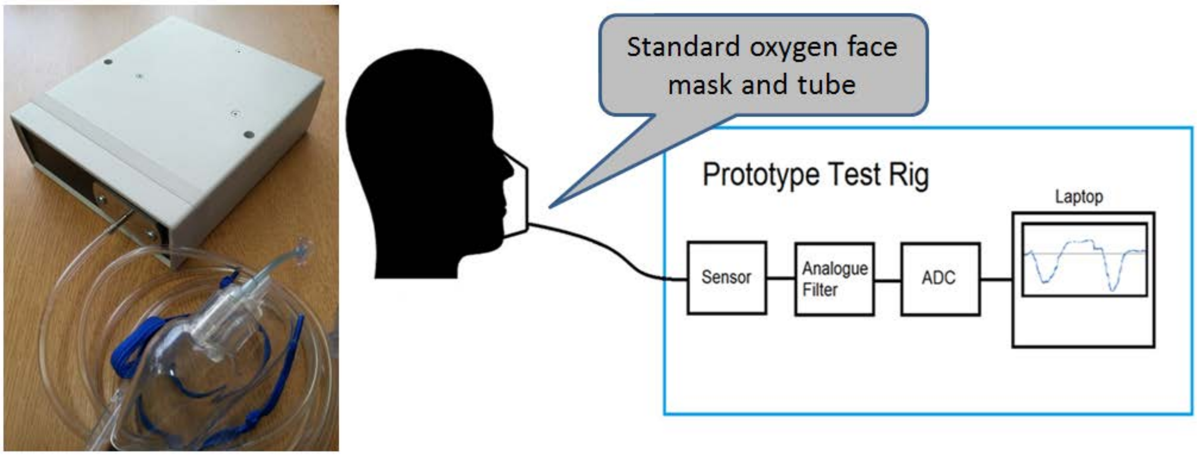

Figure 2 – (left) early prototype data collection system, showing pressure transducer and electronics housing, with breathing mask connected; (right) breath signals are filtered and digitised over a 10 second period, then displayed on the computer screen before data content analysis.

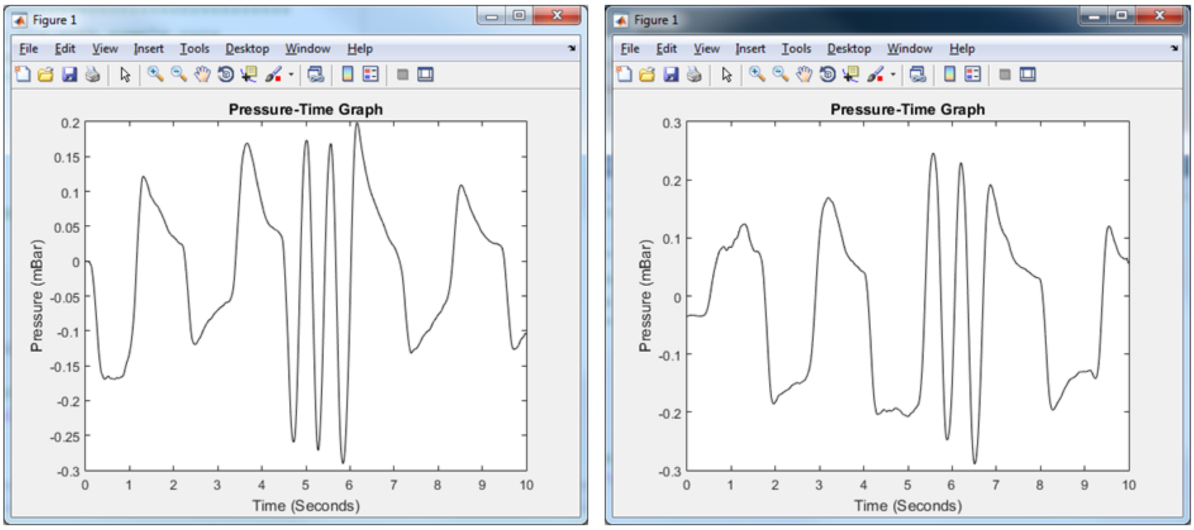

Figure 3 – breath signals collected from a pressure transducer over a 10 second period, showing the differences in signal phase between (left) inhalation at start and (right) exhalation at start.

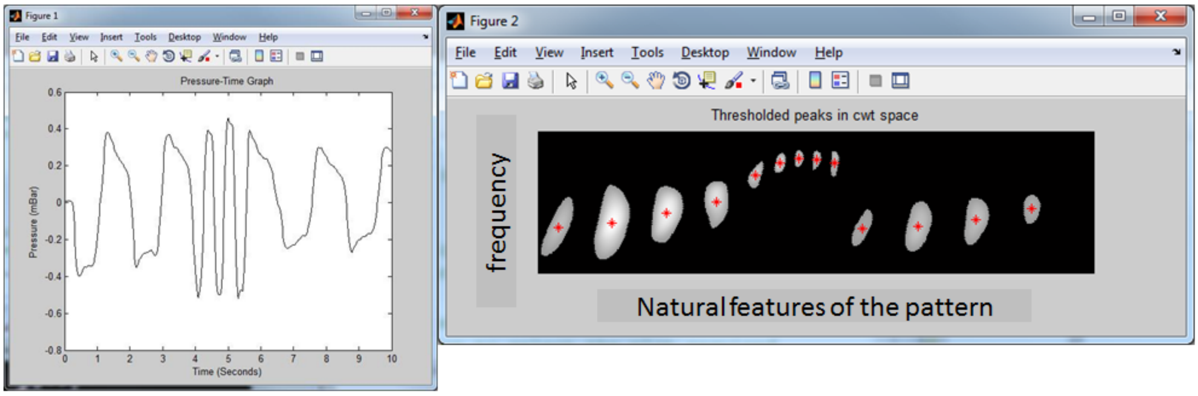

Figure 4 – (left) modulated breath pattern from pressure sensor over a 10 second period, showing variations in breathing rate; (right) data after analysis of breathing rate variations over the same time period. The white patches show how the predominant rates vary with time.