A User-centric Design of Permanent Magnetic Articulography based Assistive Speech Technology

Lam A. Cheah[1], Jie Bai[1], Jose A. Gonzalez[2], Stephen R. Ell[3], James M. Gilbert[1], Roger K. Moore[2], and Phil D. Green[2]

[1]*School of Engineering, The University of Hull, Kingston upon Hull, U.K.*

[2]*Department of Computer Science, The University of Sheffield, Sheffield, U.K.*

[3]*Hull and East Yorkshire Hospitals Trust, Castle Hill Hospital, Cottingham, U.K.*

Keywords: Silent Speech Interface, Assistive Speech Technology, Permanent Magnetic Articulography, User-centric Design.

Abstract: This paper addresses the design considerations and challenges faced in developing a wearable silent speech interface (SSI) based on Permanent Magnetic Articulography (PMA). To improve its usability, a new prototype was developed with the involvement of end users in the design process. Hence, desirable features such as appearance, portability, ease of use and light weight were incorporated into the prototype. The device showed a comparable performance with its predecessor, but has a much improved appearance, portability and hardware in terms of miniaturisation and cost.

1 INTRODUCTION

The ability to communicate through speech is crucial to humanity and plays a vital role in our social and work life. Patients whose voice box has to be removed because of throat cancer, trauma, destructive throat infections or neurological problems will inevitably lose their ability to speak. Hence, they may experience severe impact on their lives which can lead to social isolation and depression (Fagan et al., 2008). However, conventional speech restoration methods (e.g. oesophageal speech, the electrolarynx and speech valves) have limitations in terms of quality of speech and usability (Fagan et al., 2008; Gilbert et al., 2010). Furthermore, in the case of implanted speech valves, frequent valve replacement is required within a time span of 3-4 months, because of the growth of biofilm coating over time (Ell et al., 1995, 1996; Heaton and Parker, 1994).

In order to overcome these limitations, a radical alternative approach has been introduced: silent speech interfaces (SSIs). SSIs are devices that allow speech communication in the absence of audible acoustic signals. Besides their usage as communication aid for post-laryngectomy patients, SSIs can also be deployed in a quiet, noisy or acoustically challenging environment (Denby et al., 2010). Although SSIs are still in their experimental stage, there were encouraging reports over recent years. To date, there are several types of SSIs using different modalities (Denby et al., 2010). *Permanent Magnetic Articulography* (PMA) is a measuring technique that captures the magnetic field variations from a set of permanent magnets attached to the articulators (i.e. lips and tongue) during speech (Gilbert et al,. 2010). This does not provide explicit information regarding the position of the attached magnets. Rather, the measured PMA data is the combination of the magnetic field patterns associated to a particular articulatory gesture.

Despite the attractive attributes of SSIs, there are still challenges in the form of processing software (e.g. efficiency, robustness and reliable speech generation) and hardware (e.g. portability, light weight, unobtrusiveness and wearability). Preliminary discussion on the influential factors of the SSIs' implementation had been reported by Denby et al. (Denby et al., 2010), based upon criteria such as ability to operate in silence and noisy environments, usability by laryngectomee, issue of invasiveness, market readiness and cost.

The main focus of this paper is to address the hardware challenges facing the PMA-based SSI system (as opposed to the speech processing

A. Cheah L., Bai J., A. Gonzalez J., R. Ell S., M. Gilbert J., K. Moore R. and D. Green P..
A User-centric Design of Permanent Magnetic Articulography based Assistive Speech Technology.
DOI: 10.5220/0005354601090116
In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing* (BIOSIGNALS-2015), pages 109-116
ISBN: 978-989-758-069-7
Copyright © 2015 SCITEPRESS (Science and Technology Publications, Lda.)

challenges which are addressed elsewhere). A number of significant steps have been taken in order to develop a wearable system that is appropriate for everyday use. A novel embodiment comprising miniaturised sensing modules and wireless headset that is compact and comfortable is proposed.

The remainder of this paper is structured as follow. The next section briefly overviews the PMA technique and its development to date. Section 3 describes the system architectural design and the challenges, followed by the performance evaluation in section 4. The final section concludes and provides an outlook for future research.

2 RELATED WORK

The Magnetic Voice Output Communication Aid (MVOCA) is a PMA-based device developed within DiSArM (Digital Speech Recovery from Articulator Movement, www.hull.ac.uk/speech/disarm) project, aiming to restore speech communication ability for patients who have undergone surgical removal of the larynx. In principal, the MVOCA consists of an array of magnetic sensors mounted onto a lightweight headset for detection, a set of permanent magnets, four on the lips (ø1mm x 5mm), one at tongue tip (ø2mm x 4mm) and one at tongue blade (ø5mm x 1mm) as illustrated in figure 1. Information on magnets placement was described in (Gilbert et al., 2010). The magnets are temporarily attached using Histoacryl surgical tissue adhesive (Braun, Melsungen, Germany). For long term usage, these magnets will be surgically implanted. A data acquisition/control unit is used to condition the acquired measurements before transmitting them to a computer where appropriate processing and recognition algorithms are then applied.

In previous publications (Gilbert et al., 2010; Hofe et al., 2013b), the validity of the performance on isolated-word and connected digits recognition tasks using the PMA technology were shown. This was then followed by investigation into the performance across multiple speakers (Hofe et al., 2013a). A feasibility study of direct speech synthesis bypassing the intermediate recognition step was reported in (Hofe et al., 2011). More recently, extensive investigation into effectiveness of PMA data in terms discriminating the voicing, place and manner of articulation of English phones was presented in (Gonzalez et al., 2014).

So far, the tests speech experiments were carried out using the 1st generation MVOCA, which consisted of five tri-axial Honeywell HMC2003 magnetic sensors, mounted on a pair of safety glasses, as shown in figure 2. The fluctuation in magnetic field is captured on the 15 PMA channels and recorded onto a PC via ADLink DAQ-2206 analogue-to-digital converter (ADC), a PCI-based card with 16-bit linear encoding. The 2nd generation MVOCA was developed and was first used in our recent study (Gonzalez et al., 2014). Detailed description of the latest hardware will be discussed in the next section.

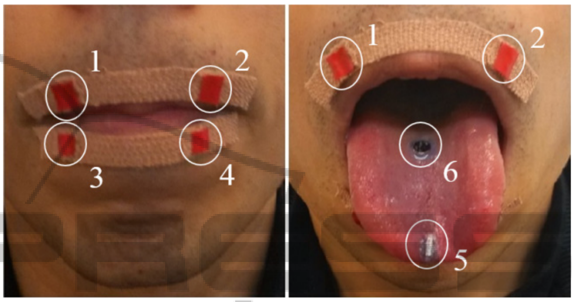

Figure 1: Placement of six magnet pellets onto lips and tongue. The magnets are cylinders with diameter and length of 1mm x 5mm for lips (pellets 1-4), 2mm x 4mm for tongue tip (pellet 5) and 5mm x 1mm for tongue blade (pellet 6).

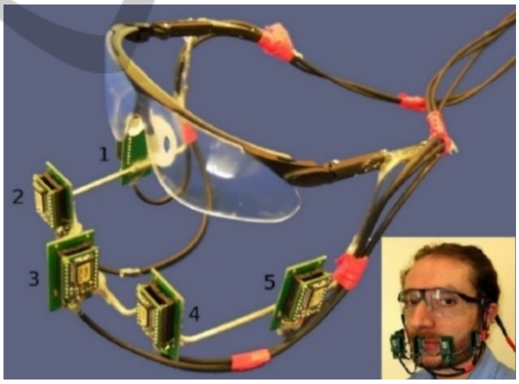

Figure 2: MVOCA headset (1st generation) - five magnetic sensors mounted on a frame that attached onto a pair of safety glasses. Appearance of device when worn by user.

3 SYSTEM ARCHITECTURE

3.1 Design Consideration

Despite encouraging performance, the earlier MVOCAs (Gilbert et al., 2010; Hofe et al., 2013a, 2013b) were not satisfactory particularly in its appearance, comfort and ergonomic factors for the user. Hence, to make the MVOCA more usable and desirable, the prototype has undergone several iterative design cycles over the past 12 months. The main focus during the development phase was to consider underexplore user-inclusive requirements by using a qualitative methodology, including informal opinion survey, focus group and user observation. These approaches were commonly used in other user-centred design studies (Bright and Conventry, 2013; Hirsch et al., 2000; Martin et al., 2006).

Based on the discussion with the panels (i.e. laryngectomees) of the focus group and data from the survey questionnaires of 50 potentials users and their families/friends, the appearance of the device is the major factor affecting the acceptability. In fact, other researches also indicated that the appearance is considered a highly desirable feature for any assistive devices (Cook and Polgar, 2008; Hirsch et al., 2000). Six possible configurations were presented in the survey, those resembling a Bluetooth earpiece or a pair of spectacles were preferred with the majority of the votes, while the device resembling a headset microphone was marginally acceptable by approximately 25% of the respondents. On the other hand, devices that might be obstructive to the mouth in anyway (in full or partially, such as the 1st generation MVOCA as illustrated in figure 2) were deemed unacceptable.

In addition, from previous focus group meetings and observation studies (participants had given their consent and the studies were approved by The University of Hull ethics committee), valuable feedback was gained and has greatly influenced the creation of a user-centric design prototype. During the prototype development, critical design questions were raised, in term of headset appearance, portability, weight, ease of use and cost.

Besides hardware appearance, the survey questionnaires also identified other desirable features, such as processing features (see table 1) and speech quality (see table 2), by their preferred ranking. As indicated in table 1, the quality of reconstructed speech is highly desirable, whereas the issue of delay between reconstructed sound and lips movement is least prioritised. In term of speech quality, this was further subdivided into the characteristics listed in table 2. Both intelligibility and naturalness of speech are considered equally desirable, but the ability to convey emotion into the reconstructed speech is least preferred. It should be noted that respondents to the survey may have had some difficulty interpreting the meaning of some of these terms since, for instance, they may not be aware of the extent of emotion present in normal speech. These non-hardware related features will not be discussed in this paper but will be addressed separately in our future work.

3.2 New MVOCA Prototype

Major hardware components of the latest MVOCA prototype are shown in figure 3. A set of four tri-axial Anisotropic Magnetoresistive (AMR) magnetic sensors (Honeywell HMC5883L), a control unit and a power source (rechargeable 7.4V Lithium Ion battery) were mounted on a customised headset. Two headsets design were developed, 1) attached onto a headband (see figure 4a), and 2) onto a pair of spectacles (see figure 4b). The headsets (excluding

Table 1: Desirable processing features.

Software feature	Description	Ranking
Speech quality	Measuring the quality of reconstructed speech (see also Table 2)	1st
Speech mode	Ability to communicate in fluent speech (ranging from isolated words to fluent speech)	2nd
Vocabulary	Size and range of words available in the database (ranging from a small context specific vocabulary to unrestricted vocabulary)	3rd
Speaking delay	Synchronisation between lips movement and synthesised voice (ranging from speaking a complete phrase before any speech output to no delay)	4th

Table 2: Desirable speech quality.

Speech quality	Description	Ranking
Intelligibility	Ability to communicate intelligibly (i.e. ranging from barely intelligible to a BBC newsreader)	1st
Naturalness	Ranging from a monotonic electronic voice to natural speech	2nd
Personification	The choice of using own or preferred voice (ranging from another appropriate voice to the user's own voice)	3rd
Ability to convey emotion	Ability to include emotions (ranging from no emotional content to full emotion content)	4th

the pair of spectacles or headband) were fabricated using rapid prototyping technology and their building material were VeroWhitePlus RGD835 and VeroBlue RGD840. A set of six Neodymium Iron Boron (NdFeB) permanent magnets are attached onto the lips and tongue as illustrated in figure 1.

Each magnetic sensor has three orthogonal sensing elements to measure the three spatial components of the magnetic field. Sensor1-3 (9 PMA channels) are used to capture magnetic field variations caused by articulatory movements and digitize it with 12-bit resolution. Whereas the sensor4 is used for background cancellation, which is for compensating the effect of earth's magnetic field thus enhancing the signal-to-noise (SNR) of the desire signals.

Figure 5 shows an operational block diagram of the 2nd generation MVOCA. Each magnetic sensor communicates to a low-power ATmega328P microcontroller (housed inside the control unit) through an I^2C interface, samples were acquired at 100 Hz. These samples (total of 12 PMA channels) are then transmitted to computer/tablet PC wirelessly via Bluetooth or USB for further processing. A bespoke graphical user interface (GUI) had been developed in the MATLAB environment and used mainly for on-line recognition testing or demonstration purposes. All necessary speech processing and recognition algorithms were embedded into the GUI and running in the background. If the acquired PMA signal correctly matched an articulation gesture from the pre-stored training dataset, thus the corresponded utterance will be identified. A text-to-speech synthesiser is used to generate a playback audio as an output for the identified utterance, via an audio device (e.g. integrated speaker of a computer).

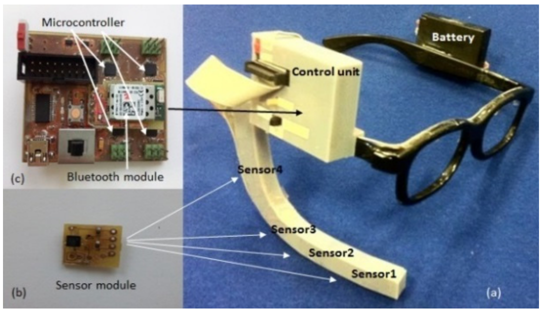

Figure 3: Overview of the 2nd generation MVOCA system, a) MVOCA headset with b) sensor modules, c) control unit and battery.

Figure 4: Two MVOCA headset designs were presented, a) mounted onto headband and b) attached onto a pair glasses. Appearances of device when worn by user.

For wireless data transmission, a class 2 Bluetooth module BTM411 (housed inside the control unit) and USB transceiver (attached on computer) are used. The MVOCA device will

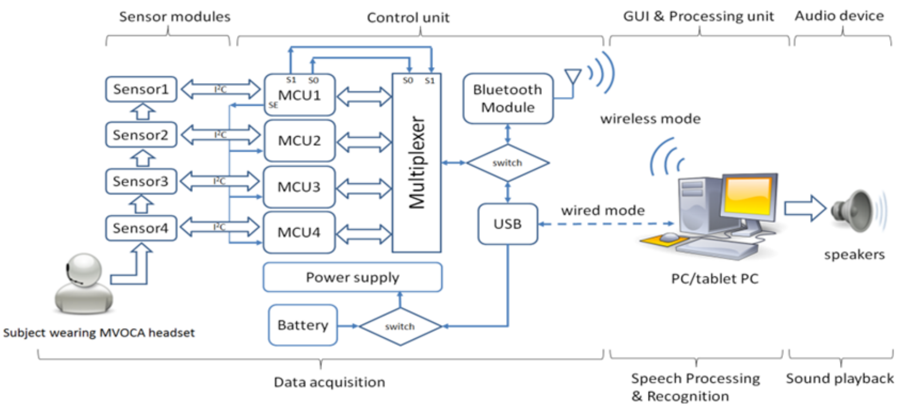

Figure 5: Simplified MVOCA (2nd generation) operational block. diagram.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

acquire its power from a battery rather than from the computer via USB (wired mode). The average power consumption of the current MVOCA prototype from a 5V (regulated from 7.4V) supply is ~104 mA, which means that it can run continuously for ~10 hours on a full charge (total 1080mAh). The battery can be removed from the headset for charging using a freestanding charger.

4 PERFORMANCE EVALUATION

4.1 Experimental Design

The data used in evaluating the latest MVOCA prototype were collected from a male native English speaker who is proficient in the usage of the MVOCA interface. As clarified previously (Hofe et al., 2013b), the MVOCA is a speaker-dependant device, i.e. all associated headset measurements and training parameters were calibrated towards particular individual. Although inter-speaker performance has proven possible (Hofe et al., 2013a), the headset design and measurements would require individually tailored for optimal performance. In particular, the headset was specifically designed according to the speaker anatomy.

To evaluate the performance of the new MVOCA prototype, a continuous speech recognition task consisting in the identification of sequences of English digits is chosen in this paper. This task was chosen because the limited size of vocabulary enables whole-word model training from relatively sparse data and also because of the simplicity of the language model involved. The algorithm used to generate the random digits sequences was the one underlying the TIDigits database of connected digits (Leonard, 1984). The longest digit sequence consists of seven individual digits. During the training, both zero and oh (the two representations of 0) were denoted as separate items.

The experimental data were collected from six independent training sessions, i.e. two sessions each one using different 2nd generation MVOCA headsets (figure 4a and 4b) and the remained two sessions using the 1st generation headset (see figure 3). Each training session consisted of five sets of data, with 4 sets (three spoken data and one mouthed data) used as training and remained mouthed data for testing. The idea is to mimic a scenario where the laryngectomy patient, where his/her has intact voice and generate training data before the laryngectomy operation. A total of 385 utterances (5 subsets) containing 1265 individual digits were recorded during each training session.

4.2 Instrumental Setup

For optimal recording performance, the experiments were conducted inside a sound-proof room, where the audio signal was recorded with a shock-mounted AKG C1000S condenser microphone and a dedicated USB sound card (Lexicon Lambda). A Matlab-based GUI was created to simultaneously record the audio signal (sampled at 48 kHz) and PMA data (sampled at 100 Hz). Since both data streams were measured from separate modality, synchronization between the two data streams was necessary to compensate for any small deviation from the ideal sampling frequencies of the analog-to-digital converters (ADC). An automatic timing alignment mechanism was used to realign both data streams by generating start-stop markers in addition to both audio and PMA data streams, in order to minimised any potential timing error.

The GUI software also provided visual prompt of the digit sequences to the speaker at regular interval of 5 seconds during the recording session. The measured PMA data were transferred to a PC via USB connection. Since the speaker's head was not restrained, large movements could potentially distort the recorded data and thus degrade the recognition performance. Hence, background cancellation was applied to compensate for any movement induced interference against the desired PMA signals.

4.3 HMM Training and Recognition

PMA data used for speech recognition was first low-pass filtered (i.e. removal 50 Hz noise) and normalized as described in (Hofe et al., 2013b). Next, the first-order time derivatives (i.e. delta parameters) were computed for each PMA frame, resulting in a feature vector of size 18. The second-order derivatives (i.e. delta-delta parameters) were not included as part of the feature vector since, as shown in our previous works (Hofe et al., 2013a, 2013b), they did not produced significant improvement in performance.

Then, the processed PMA data were used for training the speech recognizer using the HTK (Young et al., 2009). The acoustic model in the recognizer uses whole-word Hidden Markov Models (HMMs) with 25 states and 5 Gaussians per state (Hofe et al., 2013b). By no means were these parameters optimal, but the suggested parameters settings were known for their performances from our previous works (Hofe et al., 2013a, 2013b). For clarification, audio signals were not used to train the recognizer, but only the PMA data.

4.4 Recognition Performance between 1St and 2Nd Generation MVOCAs

Word and sequence accuracy results across multiple MVOCA devices are presented in figure 6 and 7. The light bars relates to the condition in which only the static PMA data is used (i.e. dimension of the feature vector is 9), whereas the darker bars refers to using both static and dynamic features (vector size of 18). We will refer to these two conditions as *Sensor* and *SensorD* features, respectively. The results reflect the mean of the data collected on two independent training sessions on each of the 1st and 2nd generation MVOCA devices.

These data were analysed independently session-by-session, and an averaged across the sessions was produced. Merging all the data from different session for recognition would seem a more attractive approach, but this might lead to inconsistent outcomes as very precise repetitive magnets placement are required on each training session. Nonetheless this could be overcome, as the magnets will be surgically implanted in the final MVOCA for long term usage. Investigations into session-independent approach on other SSIs technique were presented in (Maier-Hein et al., 2005) and (Wand and Schultz 2011).

As seen in both figure 6 and 7, it is clear that *SensorD* performed significantly better than using *Sensor* data alone. Similar trends were also reported in (Hofe et al., 2013b). Moreover, the results showed a comparable performance between the 1st and 2nd generation MVOCAs and in some cases the newer MVOCA performed slightly better. Hence, this suggests that the newer MVOCA can have better hardware features (i.e. appearance, light weight and mobility) but without compromising its recognition performance by using miniaturised components (i.e. sensors and data acquisition unit).

Figure 8 illustrates that the inclusion of mouthed data in the training dataset improves the recognition accuracy, particularly in terms of sequence recognition. In this figure, the light bars relate to mixed training data (spoken and mouthed data) and the darker ones to non-mixed training data (spoken only data). The results presented in figure 8 were trained and tested using only *SensorD* data from the 2nd generation MVOCA devices, as they provided better performance as illustrated in figure 6 and 7. Although further investigation is needed, we recognised the importance of mixing both spoken and mouthed data in any training session.

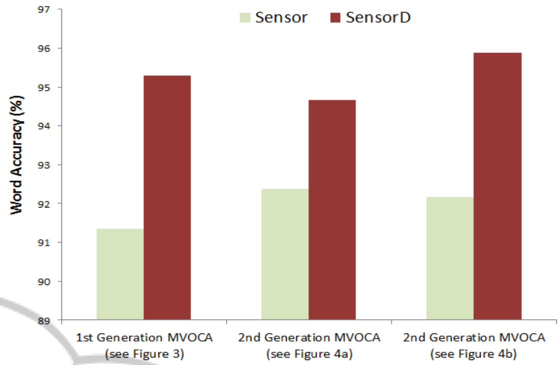

Figure 6: Comparison of word accuracy of connected digits between 1st and 2nd generation MVOCAs.

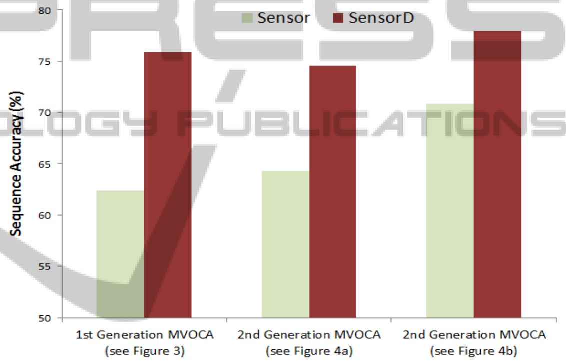

Figure 7: Comparison of sequence accuracy of connected digits between 1st and 2nd generation MVOCAs.

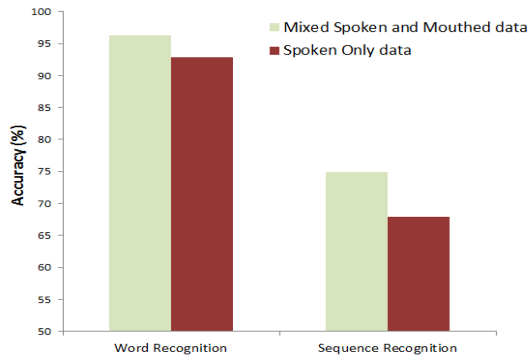

Figure 8: Comparison of training dataset (mixed or non-mixed data) in connected digits recognition rate.

4.5 Hardware Comparison between 1St and 2Nd Generation MVOCAs

The greatest challenge here is to satisfy the design objective to improve the MVOCA's appearance,

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

without compromising the device's performance. A summary of the key features of the latest MVOCA system is presented in table 3.

Table 3: Hardware specifications of the MVOCA.

Specification	Parameter
Sensor Module	
Type	Anisotropic Magnetoresistive
Dimension	12 x 12 x 3 mm^3
Sensitivity	440 LSb/gauss
Sampling rate	100 Hz/sensor
No. channels	12 (3 per sensor)
Control Unit	
Microcontroller	Low power ATmega328P
Dimension	50 x 60 x 15 mm^3
Operating voltage	5 V
Power source	Lithium Ion battery
Headset	
Material	VeroBlue/VeroWhitePlus resin
Total weight	160g (including battery & control unit)

Two versions of MVOCA headsets were designed (see figure 4a and 4b), both headsets aim to provide the desirable features such as light weight, comfort and fashionable appearance as suggested by the survey questionnaires. The current designs significantly reduced the unattractive appearance of the previous headset (see figure 2), this would improve the acceptability by the end user and ultimately improves its usage.

Significant improvements were also made in term of the hardware miniaturisations and portability. The previous MVOCA relied on a PCI-based data acquisition card, thus restricted it to a desktop PC/workstation which is highly immobile and bulky. Although the magnetic sensors HMC2003 are high precision sensors, they are significant larger in size (24x45x10 mm^3) and required higher operation voltage (i.e. 12V), thus making them non-power efficient.

In the current prototype, magnetic sensors HMC5883L were chosen because of their compactness, low operation voltage, low cost and wide sensitivity range. As for signal conditioning, low-powered microcontrollers were used. By utilising a Bluetooth modules and a tablet PC (i.e. mobile processing unit), the current MVOCA will be highly portable and practical for everyday use. In addition, the cost of the prototyping is relatively low, as the MVOCA only utilised commercial off-the-shelf (COTS) components. Moreover, by shrinking the size of electronics, this inevitably reduced the overall weight of the headset, and making it more appealing as a wearable assistive speech technology.

The downside of the new MVOCA prototype would be the omission of higher precision components (i.e. magnetic sensors) used in previous prototype, reduction in numbers of sensors and the use of a lower sampling rate. However, from the results presented in figure 6 and 7, these would suggest the concerns would be irrelevant as the performances are comparable and slightly better in some cases when using the 2^{nd} generation MVOCA. There could be a couple of good explanations for that. Firstly, the articulation movements during speech is slow and therefore a lower sampling rate (i.e. 100 Hz) might be sufficient. Secondly, reduction in number of sensors was possible because there were excess of information available from previous MVOCA, thus some sensors can be made redundant.

5 CONCLUSIONS

The preliminary evaluation of the new MVOCA prototype shows comparable recognition performances to the previous system, but providing much desirable hardware improvements such as portability, hardware miniaturisation, desirable appearance and lower cost. Nevertheless, there are still many challenges ahead before MVOCA can be practically operated outside laboratory environments on a day-to-day basis.

Encouraged by the results obtained so far, extensive work is needed to create a viable wearable assistive communication aid. Therefore, future works include enhancing overall MVOCA appearance, reducing power consumption and implementing real-time features (e.g. reducing processing time delay and utilising on-line endpointing/segmentation). In addition, investigation into the work on speech synthesis from PMA data had started and preliminary results were very encouraging.

ACKNOWLEDGEMENTS

The report is an independent research funded by the National Institute for Health Research (NIHR)'s Invention for Innovation Programme. The views stated are those of the authors and not necessary reflecting the thoughts of the sponsor.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

REFERENCES

Bright, A. K., and Conventry, L., 2013. Assistive technology for older adults: psychological and socio-emotional design requirements, *Proceedings of the 6^{th} International Conference on PErvaesive Technologies Related to Assistive Environments*, pages 1-4, Rhodes, Greece.

Cook, A. M., and Polgar, J. M., 2008. Assistive Technologies: Principles and Practice, Mosby-Year Book, 3^{rd} edition.

Denby, B., Schultz, T., Honda, K., Hueber, T., Gilbert, J. M., and Brumberg, J. S., 2010. Silent speech interfaces. *Speech Communication*, 52(4):270-287.

Ell, S. R., Mitchell, A. J., and Parker, A. J., 1995. Microbial colonisation of the Groningen speaking valve and its relationship to valve failure. *Clinical Otolaryngol*, 20:555-556.

Ell, S. R., Mitchell, A. J., Clegg, R. T., and Parker, A. J., 1996. Candida: the cancer of silastic. *Journal of Laryngol Otol*, 110(3):240-242.

Fagan, M. J., Ell, S. R., Gilbert, J. M., Sarrazin, E., and Chapman, P. M., 2008. Development of a (silent) speech recognition system for patients following laryngectomy. *Medical Engineering & Physics*, 30(4):419-425.

Gilbert, J. M., Rybchenko, S. I., Hofe, R., Ell, S. R., Fagan, M. J., Moore, R. K. and Green, P. D., 2010. Isolated word recognition of silent speech using magnetic implants and sensors. *Medical Engineering & Physics*, 32(10):1189-1197.

Gonzalez, J. A., Cheah, L. A., Bai, J., Ell, S. R., Gilbert, J. M., Moore, R. K., and Green, P. D., 2014. Analysis of phonetic similarity in a silent speech interface based on permanent magnetic articulography. *Proceedings of 15^{th} INTERSPEECH*, pages 1018-1022, Singapore.

Heaton, J. M., and Parker, A. J., 1994. Indwelling tracheа-oesophageal voice prostheses post-laryngectomy in Sheffield, UK: a 6-year review. *Acta Oto-Laryngol*, 114:675-678.

Hirsch, T., Forlizzi, J., Goetz, J., Stoback, J., and Kurtx, C., 2000. The ELDer project: Social and emotional factors in the design of eldercare technologies. *Proceedings on the 2000 conference of Universal Usability*, pages 72-79, Arlington, USA.

Hofe, R., Bai, J., Cheah, L. A., Ell, S. R., Gilbert, J. M., Moore, R. K., and Green, P. D., 2013a. Performance of the MVOCA silent speech interface across multiple speakers. *Proceedings of 14^{th} INTERSPEECH*, pages 1140-1143, Lyon, France.

Hofe, R., Ell, S. R., Fagan, M. J., Gilbert, J. M., Green, P. D., Moore, R. K., and Rybchenko, S. I., 2013b. Small-vocabulary speech recognition using silent speech interface based on magnetic sensing. *Speech Communication*, 55(1):22-32.

Hofe, R., Ell, S. R., Fagan, M. J., Gilbert, J. M., Green, P. D., Moore, R. K., and Rybchenko, S. I., 2011. Speech synthesis parameters generation for the assistive silent speech interface MVOCA. *Proceedings of 12^{th} INTERSPEECH*, pages 3009-3012, Florence, Italy.

Leonard, R. G., 1984. A database for speaker-independent digit recognition. *International Conference on Acoustic, Speech and Signal Processing*, pages 328-331, San Diego, USA.

Maier-Hein, L., Metze, F., Schultz, T., and Waibel, A., 2005. Session independent non-audible speech recognition using surface electromyography. *Automatic Speech Recognition and Understanding Workshop*, pages 331-336, Cancun, Mexico.

Martin, J. L., Murphy, E., Crowe, J. A., and Norris, B. J., 2006. Capturing user requirements in medical devices development: the role of ergonomics, *Physiological Measurement*, 27(8):R49-R62.

Wand, M., and Schultz, T., 2011. Session-independent EMG-based speech recognition. *4^{th} International Conference on Bio-inspired Systems and Signal Processing*, pages 295-300, Rome, Italy.

Young, S., Everman, G., Gales, M., Hain, T., Kershaw, D., Liu, X., Moore, G., Odell, J., Ollason, D., Povery, D., Valtchev, V., and Woodland, P., 2009. The HTK Book (for HTK Version 3.4).