The final publication is available at Springer via https://link.springer.com/book/10.1007/978-3-319-27707-3.

Integrating User-Centred Design in the Development of a Silent Speech Interface based on Permanent Magnetic Articulography

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

[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.
{l.cheah, j.m.gilbert, j.bai, m.j.fagan, s.i.rynchenko}@hull.ac.uk, {j.gonzalez, r.k.moore, p.green}@sheffield.ac.uk, srell@doctors.org.uk

Abstract. A new wearable silent speech interface (SSI) based on Permanent Magnetic Articulography (PMA) 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 integrated into the prototype. The aim of this paper is to address the challenges faced and the design considerations addressed during the development. Evaluation on both hardware and speech recognition performances are presented here. The new prototype shows a comparable performance with its predecessor in terms of speech recognition accuracy (i.e. ~95% of word accuracy and ~75% of sequence accuracy), but significantly improved appearance, portability and hardware features in terms of miniaturization and cost.

Keywords: Assistive Speech Technology, User-Centred Design, Silent Speech Interface, Permanent Magnetic Articulography, Magnetic Sensors

1 Introduction

Speech is an important part of human communication and plays a vital role in our social and work life. There are many situations in which people wish to communicate through speech but where it is either impossible (i.e. medical condition) or not desirable (i.e. communicating in private or in noisy environment). 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. Therefore, they may experience a severe impact on their lives which can lead to social isolation and depression [1]. Conventional speech restoration methods after laryngectomy (e.g. oesophageal speech, the electrolarynx and speech valves) have limitations in terms of quality of speech and usability [1,2]. Moreover, 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 [3,4,5]. To address these shortcomings, a novel approach has been introduced: silent speech interfaces (SSIs).

SSIs are devices that enable speech communication in the absence of audible acoustic signals. To do that, SSIs exploit other non-acoustic information generated during the speech production process. These alternative sources can range from brain activity to articulator movements. To extract these forms of information several types of SSIs using different modalities have been proposed so far [6]. Permanent Magnetic Articulography (PMA) is a type of SSI and it is based on sensing the magnetic field variations from a set of permanent magnets attached to the articulators (i.e. lips and tongue) during speech [1,2]. Contrary to other similar SSIs such as Electromagnetic Articulography (EMA), PMA does not provide explicit information regarding the position of the attached magnets. Instead, the measured PMA data is the summation of the magnetic field patterns associated to a particular articulatory gesture. As will be shown later, this is not a limitation of PMA as long as captured articulator data is used for pattern recognition (e.g. speech recognition). In this case, pattern recognition techniques can be employed to recognize the PMA patterns associated to the particular speech sounds.

Although there are obvious advantageous in using SSIs, there are still challenges in the form of the processing software (e.g. efficiency, robustness and reliable speech generation) and hardware (e.g. portability, light weight, unobtrusiveness and wearability). Preliminary investigation on the influential factors of the SSIs' implementation had been presented in [6], based upon criteria such as ability to operate in silence and noisy environments, usability by laryngectomees, issues of invasiveness, market readiness and cost. The focus of this paper is on the hardware challenges facing the PMA-based SSI system. 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 miniaturized sensing modules and a wireless headset that is compact and comfortable is proposed in this work.

The rest of this paper is organized as follow. The next section overviews the PMA technique and its development to date. Section 3 describes the design of the 2nd generation system and the associated challenges, followed by the performance evaluation in section 4. The final section concludes and provides an outlook for future research.

2 Overview on PMA-based SSI

A PMA-based device, the Magnetic Voice Output Communication Aid (MVOCA), is developed within the 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 a nutshell, the current MVOCA device consists of multiple magnetic sensors mounted onto a lightweight headset for detection, a set of permanent magnets, four on the lips (ø1mm × 5mm), one at tongue tip (ø2mm × 4mm) and one at tongue blade (ø5mm × 1mm) as illustrated in Fig. 1. Information on magnet placement was described in [2]. These magnets are temporarily attached using Histoacryl surgical tissue adhesive (Braun, Melsungen, Germany). Eventually, these magnets will be surgically implanted for long term usage. The acquired measurements are pre-conditioned by the control unit prior to further signal processing.

To date, all the 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 Fig. 2. The fluctuation in the 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. Before describing the 2nd generation MVOCA device in next section, related background works on PMA are briefly outlined as follow:

- In earlier work [2,7], the viability of isolated-word and connected digits recognition tasks using the PMA technology were presented.
- Investigation into the performance across multiple speakers conducted in [8].
- A feasibility study of direct speech synthesis bypassing the intermediate recognition step was reported in [9].
- 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 [10].

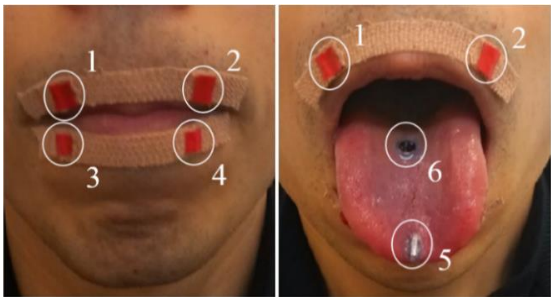

Fig. 1. Placement of six magnets with diameter and length of 1mm × 5mm for lips (pellets 1-4), 2mm × 4mm for tongue tip (pellet 5) and 5mm × 1mm for tongue blade (pellet 6).

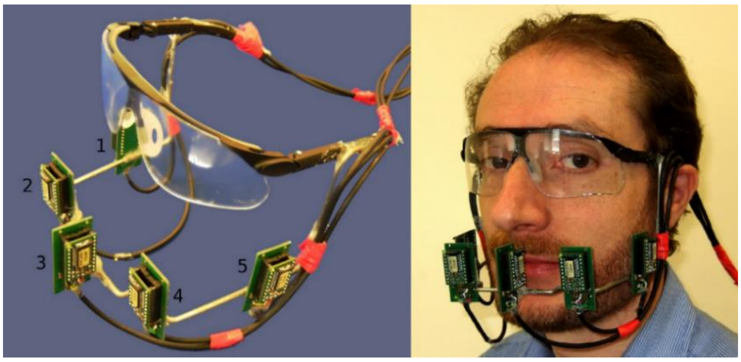

Fig. 2. MVOCA headset (1st generation) - five magnetic sensors mounted on a frame that attached onto a pair of safety glasses. Appearance of the device worn by user.

3 System Description

3.1 Design Challenges and Considerations

In order, to make the MVOCA device more usable and desirable, the 1st generation prototype has undergone several design cycles over the last 12 months. This is because the earlier MVOCAs [2,7,8] were not satisfactory, particularly in their appearance, comfort and ergonomic factors for the user, despite encouraging performance. The focus of the development task 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 [11,12,13].

Through 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 was seen as the major factor affecting acceptability. In fact, other researches also indicated that unobtrusive appearance is considered a highly desirable feature for any assistive device [12,14]. Six possible configurations were presented in the survey. Those resembling a Bluetooth earpiece or a pair of spectacles were preferred by the majority of the potential users, while the device resembling a headset microphone was marginally acceptable by approximately 25% of the respondents. On the other hand, devices that might obstruct the view of the mouth in anyway (in full or partially, such as the 1st generation MVOCA as illustrated in Fig. 2) were deemed unacceptable. Moreover, through 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-centred design prototype. Critical design questions were raised during prototype development, in term of headset appearance, portability, weight, ease of use and cost.

Table 1. Desirable software features.

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

Table 2. Desirable speech qualities.

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

In addition, the survey questionnaires also identified other desirable features, such as software 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 rated, whereas the issue of delay between reconstructed sound and lips movement is least prioritized. 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 important, 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. The non-hardware related features will not be discussed in this paper but will be addressed separately in our future work.

3.2 New MVOCA Device

Based on the information gathered from the potential users (as presented in section 3.1), a new prototype has been developed. Key components of the 2nd generation MVOCA prototype consists 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). These components are mounted on a customized headset, as illustrated in Fig. 3. Two headsets design were developed: 1) attached onto a headband (see Fig. 4a), and 2) onto a pair of spectacles (see Fig. 4b). The headsets (excluding the pair of spectacles or headband) were fabricated using rapid prototyping technology and their building materials 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 Fig. 1. Each magnetic sensor has three orthogonal sensing elements to measure the three spatial components of the magnetic field. Sensor1-3 (a total of 9 channels) are used to capture magnetic field variations caused by articulatory movements and digitize it with 12-bit resolution. Sensor4 on the other hand is used for background cancellation that is for compensating the effect of earth's magnetic field on the articulography signals captured by other sensors, in order to enhance the signal-to-noise (SNR) of the articulatory signals.

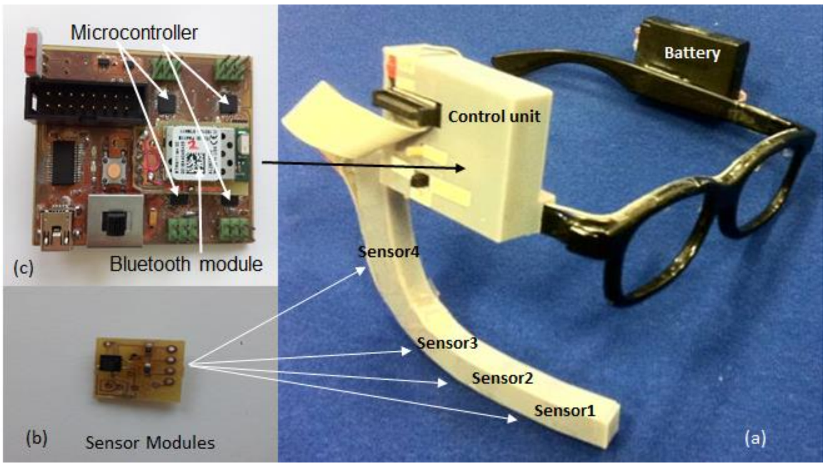

Fig. 3. Overview of the 2nd generation MVOCA system, a) MVOCA headset with b) sensor modules, c) control unit and battery.

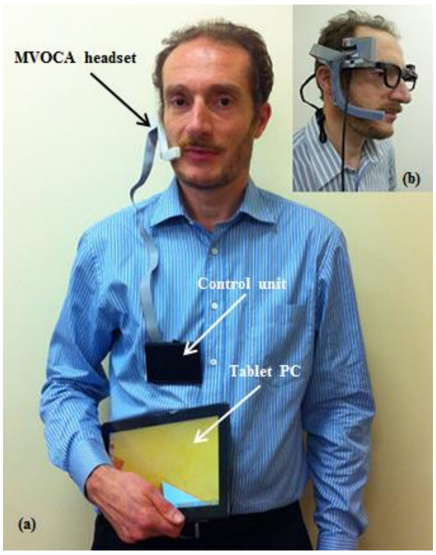

Fig. 4. Two MVOCA headset designs, a) mounted on a headband and b) attached on a pair of glasses. Appearance of the devices when worn by a user.

An operational block diagram of the 2nd generation MVOCA is shown in Fig. 5. Each magnetic sensor communicates to a low-power ATmega328P microcontroller (housed inside the control unit) through an I^2C interface, and a handful of control signals (i.e. SE, S0, and S1) are used in managing the acquisition process where samples are acquired at 100 Hz for each channel. These samples (total of 12 PMA channels) are then transmitted to a computer/tablet PC wirelessly via Bluetooth or via a USB connection for further processing. A bespoke graphical user interface (GUI) has 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 synthesizer 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).

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

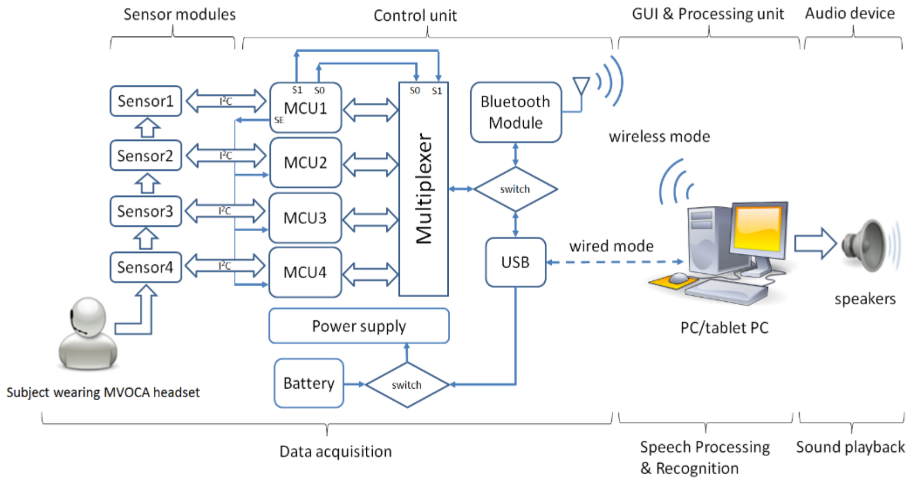

Fig. 5. Simplified MVOCA (2nd generation) operational block diagram.

4 Experiments and Results

4.1 Test Speaker

As stated previously [7], the current MVOCA is a speaker-dependent device, i.e. all associated headset measurements and training parameters were calibrated towards particular individual. In this work, the data used for evaluating the latest MVOCA prototype were collected from a male native English speaker who is proficient in the usage of the MVOCA interface. Although inter-speaker performance has proven possible [8] the headset design and measurements would require individually tailored for optimal performance. In particular, the headset was specifically designed according to the speaker's anatomy.

4.2 Data Recording

A continuous speech recognition task consisting of the identification of sequences of English digits is chosen in this work to evaluate the performance of the latest MVOCA prototype. This 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 [15]. 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 data used for training the speech recognizer were collected in six independent sessions, i.e. two sessions using each of the different 2nd generation MVOCA headsets (Fig. 4a and Fig. 4b) and the remained two sessions using the 1st generation headset (see Fig. 3). Each training session consisted of a total of 385 utterances containing 1265 individual digits. Furthermore, within each session, five different datasets were recorded: four of them (three spoken datasets and one mouthed dataset) were used for training and the remained mouthed dataset was used for testing purpose. The reason behind this configuration is to try to mimic a realistic scenario where the voice of the patient is recorded before the operation happens for personalizing the speech synthesizer, while after the operation only articulography PMA (mouthed) data can be obtained.

4.3 Experimental Setup

To achieve optimal recording performance, all experiments in this paper 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 provide visual prompt of the digit sequences to the speaker at regular interval of 5 seconds during the recording session. The GUI also used provides simultaneous recording of both audio signal (sampled at 48 kHz) and PMA data (sampled at 100 Hz) as illustrated in Fig. 6.

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). To do that 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. 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.

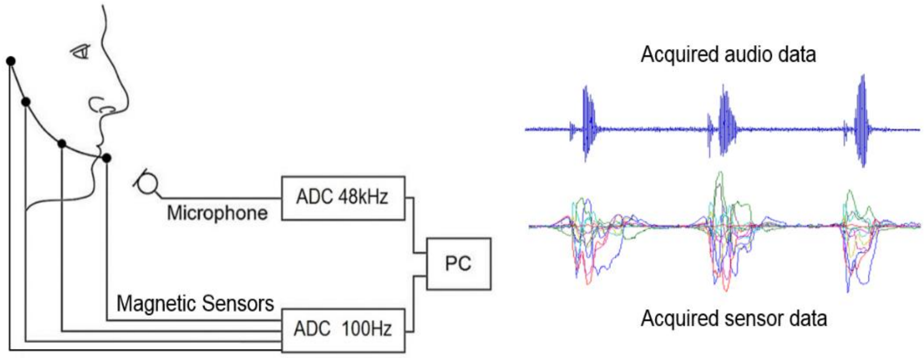

Fig. 6. Acoustic and PMA data streams were recorded in parallel into PC via a bespoke GUI. Both data streams were then synchronized prior to pre-processing.

4.4 HMM Training and Recognition

The acquired PMA data used for speech recognition was first low-pass filtered (i.e. removal of 50 Hz noise) and normalized as described in [7]. Two different conditions (i.e. *Sensor* and *SensorD*) were computed in connected digits recognition experiments, and they relate to a specific configuration of the data used for model training and testing:

- *Sensor*: training and testing directly on the 9 channels of PMA data.
- *SensorD*: as above, plus the first time derivatives (related to articulator velocity, *D* stands for “delta”).

An overview of the two conditions is presented in Table 3. 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 [7,8], they did not produced significant improvement in performance.

Table 3. Vector sizes for different experimental conditions.

Condition	Original	1st delta	Vector size
Sensor	√		9
SensorD	√	√	18

The processed PMA data were then used for training the speech recognizer using HTK [16]. The acoustic model in the recognizer uses whole-word Hidden Markov Models (HMMs) [17] with 25 states and 5 Gaussians per state [7]. These parameters were not optimal, but the suggested parameters settings were known for their performances from our previous works [7,8]. For clarification, audio signals were not used to train the recognizer, but only the PMA data.

4.5 Performance between 1st and 2nd Generation MVOCAs

Both word and sequence accuracy results across multiple MVOCA devices are presented in Fig. 7. The results reflect the averaged value of the data (i.e. *Sensor* and *SensorD*) collected on two independent training sessions on each of the 1st and 2nd generation MVOCA devices. The data were analyzed independently session-by-session, and the recognition rates averaged across the sessions. Merging all the data from different sessions 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 [18,19].

As seen in Fig. 7, it is obvious that *SensorD* performs significantly better than using *Sensor* data alone. Similar trends were also reported in [7]. Moreover, the results showed a comparable performance between the 1st and 2nd generation MVOCAs. Hence, this suggests that the newer MVOCA can have better hardware features (i.e. appearance, light weight and portability) but without compromising its recognition performance by using miniaturized components (i.e. sensors and data acquisition unit).

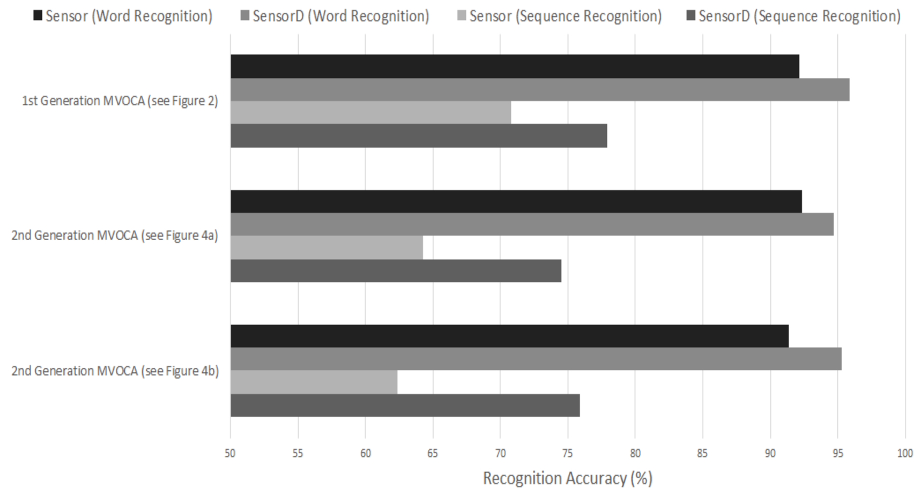

Fig. 7. . Comparison of word and sequence accuracies of connected digits between 1st and 2nd generation MVOCAs.

Fig. 8 illustrates that the inclusion of mouthed data in the training dataset improves the recognition accuracy, particularly in terms of sequence recognition. A comparison between using mixed data (spoken and mouthed data) and non-mixed data (spoken only data) as part of training dataset was investigated. The darker bars relate to mixed training data (spoken and mouthed data) and the light ones to non-mixed training data (spoken only data). The results presented in Fig. 8 were trained and tested using only

SensorD data from the 2nd generation MVOCA devices, as they provided better performance as illustrated in Fig. 7. Although further investigation is needed, we recognized the importance of mixing both spoken and mouthed data in any training session.

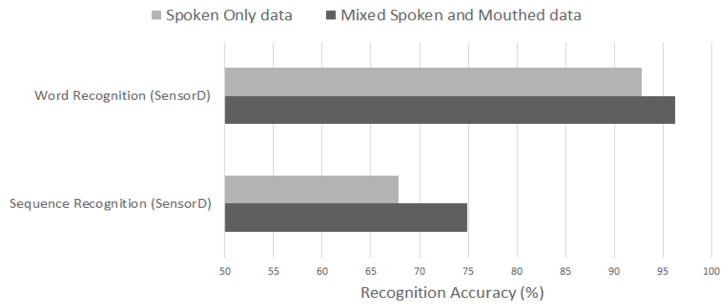

Fig. 8. Comparison of training dataset (mixed or non-mixed data) used in the recognition of connected digits.

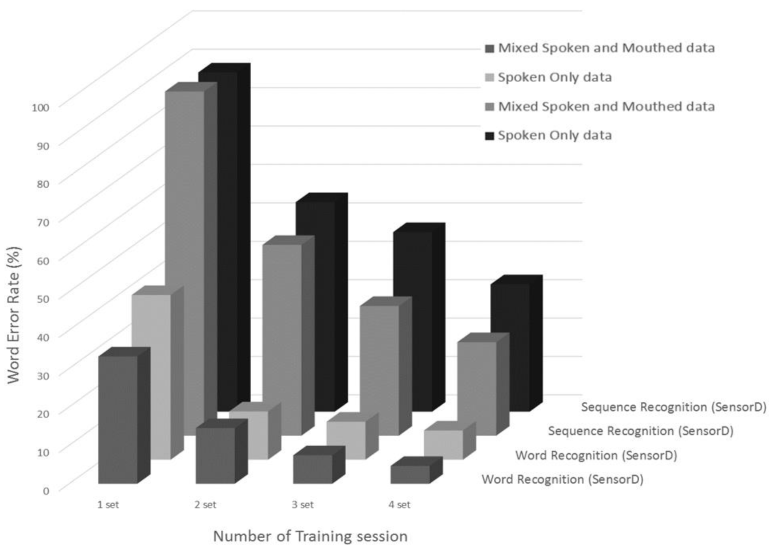

Fig. 9. Decrease in word error rate (WER) with the increase in training sessions.

So far, the results in Fig. 7 and Fig. 8 suggest that a *SensorD* data trained using a mixture of spoken and mouthed data generally performed better. A follow up test was conducted to explore the relationship between quantity of training data and the recognition performances (i.e. word and sequence recognition). Fig. 9 illustrates that an increased number of training sessions yields an improvement in the performance in both word and sequence recognitions through the reduction of in word error rate (WER). Fewer training sessions (i.e. ≥ 2 sets) is needed to achieve reasonable performance in word recognition as compared to sequence recognition, whereas it appears significantly more training sets would be required to achieve similar sequence recognition performance. It also appears that even for word recognition, the inclusion of further training data sets could reduce the WER further. The training sessions were not extended because of the speaker fatigue and increased the likelihood of the magnets becoming detached.

4.6 Hardware Comparison between 1st and 2nd Generation MVOCAs

So far, the challenge is to satisfy the design objective to improve the MVOCA's appearance, without compromising the device's performance. A summary of the key features of the latest MVOCA system is presented in Table 4. Two versions of MVOCA headsets were designed (see Fig. 4), both headsets aim to provide the desirable features such as light weight, comfort and unobtrusive appearance as suggested by the survey questionnaires. The current designs significantly reduces the unattractive appearance of the previous headset (see Fig. 2), thus this would improve the acceptability to the end user and ultimately improves its usage.

Table 4. Hardware specifications of the 2nd generation MVOCA.

	Specification	Parameter
Sensor Modules	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
	Transceiver	Bluetooth/USB
Headset	Material	VeroBlue/VeroWhitePlus resin
	Total weight	160g (including battery & control unit)

In addition, significant improvements were made in term of the hardware miniaturizations and portability, as previous generation 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 (24×45×10 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 utilizing 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 utilized commercial off-the-shelf (COTS) components. Moreover, by shrinking the size of electronics, this inevitably reduces the overall weight of the headset, and making it more appealing as a wearable assistive speech technology.

On the other hand, this would mean the omission of higher precision components (i.e. magnetic sensors) used in the previous prototype, a reduction in the numbers of sensors and the use of a lower sampling rate. However, from the results presented in Fig. 7, these concerns would appear to be irrelevant as the performances are comparable between 1st and 2nd generation MVOCAs. This could be that the articulator movements during speech are slow and therefore a lower sampling rate (i.e. 100 Hz) might be sufficient. In addition, reduction in the number of sensors was possible because there were excess of information available from previous MVOCA, thus some sensors can be made redundant.

5 Conclusion

The preliminary evaluation of the new MVOCA prototype shows comparable recognition performances to the previous system, but providing much more desirable hardware features such as portability, hardware miniaturization, improved appearance and lower cost. Nonetheless, 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. Potential future works may include enhancing overall MVOCA appearance, reducing power consumption and implementing real-time features (i.e. reducing latency in processing and decision making). On the other hand, to address the desirable features on speech quality as discussed in section 3, investigation work on speech synthesis (similar to the work in [9]) from PMA data has started and preliminary results obtained are very encouraging.

Acknowledgements. The work 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.

References

1. Fagan, M. J., Ell, S. R., Gilbert, J. M., Sarrazin, E., and Chapman, P. M.: Development of a (silent) speech recognition system for patients following laryngectomy. Medical Engineering & Physics, 30(4), 419-425 (2008).
2. Gilbert, J. M., Rybchenko, S. I., Hofe, R., Ell, S. R., Fagan, M. J., Moore, R. K. and Green, P. D.: Isolated word recognition of silent speech using magnetic implants and sensors. Medical Engineering & Physics, 32(10), 1189-1197 (2010).
3. Ell, S. R., Mitchell, A. J., and Parker, A. J.: Microbial colonisation of the Groningen speaking valve and its relationship to valve failure. Clinical Otolaryngology, 20, 555-556 (1995).

4. Ell, S. R., Mitchell, A. J., Clegg, R. T., and Parker, A. J.: Candida: the cancer of silastic. Journal of Laryngology Otology, 110(3), 240-242 (1996).
5. Heaton, J. M., and Parker, A. J.: Indwelling trachea-oesophageal voice prostheses post-laryngectomy in Sheffield, UK: a 6-year review. Acta Oto-Laryngologica, 114, 675-678 (1994).
6. Denby, B., Schultz, T., Honda, K., Hueber, T., Gilbert, J. M., and Brumberg, J. S.: Silent speech interfaces. Speech Communication, 52(4), 270-287 (2010).
7. Hofe, R., Ell, S. R., Fagan, M. J., Gilbert, J. M., Green, P. D., Moore, R. K., and Rybchenko, S. I.: Small-vocabulary speech recognition using silent speech interface based on magnetic sensing. Speech Communication, 55(1), 22-32 (2013).
8. Hofe, R., Bai, J., Cheah, L. A., Ell, S. R., Gilbert, J. M., Moore, R. K., and Green, P. D.: Performance of the MVOCA silent speech interface across multiple speakers. In: Proceedings of the 14th INTERSPEECH, pp. 1140-1143, Lyon, France (2013).
9. Hofe, R., Ell, S. R., Fagan, M. J., Gilbert, J. M., Green, P. D., Moore, R. K., and Rybchenko, S. I.: Speech synthesis parameters generation for the assistive silent speech interface MVOCA. In: Proceedings of the 12th INTERSPEECH, pp. 3009-3012, Florence, Italy (2011).
10. Gonzalez, J. A., Cheah, L. A., Bai, J., Ell, S. R., Gilbert, J. M., Moore, R. K., and Green, P. D.: Analysis of phonetic similarity in a silent speech interface based on permanent magnetic articulography. In: Proceedings of the 15th INTERSPEECH, pp. 1018-1022, Singapore (2014).
11. Bright, A. K., and Conventry, L.: Assistive technology for older adults: psychological and socio-emotional design requirements, In: Proceedings of the 6th International Conference on PErvaesive Technologies Related to Assistive Environments, pp. 1-4, Rhodes, Greece (2013).
12. Hirsch, T., Forlizzi, J., Goetz, J., Stoback, J., and Kurtx, C.: The ELDer project: Social and emotional factors in the design of eldercare technologies. In: Proceedings on the 2000 conference of Universal Usability, pp. 72-79, Arlington, USA (2000).
13. Martin, J. L., Murphy, E., Crowe, J. A., and Norris, B. J.: Capturing user requirements in medical devices development: the role of ergonomics. Physiological Measurement, 27(8), R49-R62 (2006).
14. Cook, A. M., and Polgar, J. M.: Assistive Technologies: Principles and Practice, 3rd edn. Mosby, London (2008).
15. Leonard, R. G.: A database for speaker-independent digit recognition. In: International Conference on Acoustic, Speech and Signal Processing, pp. 328-331, San Diego, USA (1984).
16. 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.: The HTK Book (for HTK Version 3.4.1), Cambridge University Press, Cambridge (2009).
17. Rabiner, L.R.: A tutorial on Hidden Markov Models and selected applications in speech recognition. Proceedings of the IEEE, 77, 257-286 (1989).
18. Maier-Hein, L., Metze, F., Schultz, T., and Waibel, A.: Session independent non-audible speech recognition using surface electromyography. In: Automatic Speech Recognition and Understanding Workshop, pp. 331-336, Cancun, Mexico (2005).
19. Wand, M., and Schultz, T.: Session-independent EMG-based speech recognition. In: Proceedings of the 4th International Conference on Bio-inspired Systems and Signal Processing, pp. 295-300, Rome, Italy (2011).