Towards an Intraoral-based Silent Speech Restoration System for Post-Laryngectomy Voice Replacement

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

[1]School of Engineering, University of Hull, Kingston upon Hull, U.K.
`{l.cheah, j.m.gilbert, j.bai}@hull.ac.uk`
[2]Department of Computer Science, University of Sheffield, Sheffield, U.K.
`{j.gonzalez, p.green, r.k.moore}@sheffield.ac.uk,`
[3]Hull and East Yorkshire Hospitals Trust, Castle Hill Hospital, Cottingham, U.K.
`srell@doctors.org.uk`

Abstract. Silent Speech Interfaces (SSIs) are alternative assistive speech technologies that are capable of restoring speech communication for those individuals who have lost their voice due to laryngectomy or diseases affecting the vocal cords. However, many of these SSIs are still deemed as impractical due to a high degree of intrusiveness and discomfort, hence limiting their transition to outside of the laboratory environment. We aim to address the hardware challenges faced in developing a practical SSI for post-laryngectomy speech rehabilitation. A new Permanent Magnet Articulography (PMA) system is presented which fits within the palatal cavity of the user's mouth, giving unobtrusive appearance and high portability. The prototype is comprised of a miniaturized circuit constructed using commercial off-the-shelf (COTS) components and is implemented in the form of a dental retainer, which is mounted under roof of the user's mouth and firmly clasps onto the upper teeth. Preliminary evaluation via speech recognition experiments demonstrates that the intraoral prototype achieves reasonable word recognition accuracy and is comparable to the external PMA version. Moreover, the intraoral design is expected to improve on its stability and robustness, with a much improved appearance since it can be completely hidden inside the user's mouth.

Keywords: Silent speech interface, Assistive technology, Wireless intraoral device, Permanent magnet articulography, Magnetoresistive sensors

1 Introduction

Speech is perhaps the most convenient and natural form of human communication. Patients who have had a laryngectomy (e.g. surgical removal of larynx as part of treatment for cancer or other diseases affected the vocal cords) lose their voices and often struggle with their daily communication. Hence, they may experience severe impact on their lives which can lead to social isolation, loss of identity and depression [1, 2]. However, there are currently only a limited number of post-laryngectomy voice

adfa, p. 1, 2011.

restoration methods available for these individuals: esophageal speech, the electrolarynx and speech valves. Unfortunately, these methods are often limited by their usability and/or the abnormal voice produced, which may be hard to understand for listeners [1, 3-4]. On the other hand, typing-based augmented and alternative communication (AAC) devices are limited by slow manual text input [5]. Although some improvements have been achieved in term of the voice quality of the electrolarynx and esophageal speech [6, 7], emerging assistive technologies (ATs) such as silent speech interfaces (SSIs) have shown promising potential in recent years as an alternate solution.

SSIs are devices that enable speech communication to take place in the absence of audible acoustic signals [8]. To date, a number of SSIs have been proposed in an attempt to extract non-acoustic information generated during speech production and reproduce audible speech using different sensing modalities, such as measuring electrical activities of the brain [9-11] or the articulator muscles [12-14], or by capturing movements of the speech articulators themselves [3, 5, 8, 15-19]. A comprehensive summary on different SSIs technologies were presented in [8]. Because of their unique feature, SSIs can also be deployed in acoustically challenging environment or where privacy/confidentially is desirable, and not limited to its use as a communication aid for speech impaired individuals.

Despite the attractive attributes of SSIs, there are still challenges in the form of hardware (e.g. portability, lightweight, unobtrusiveness and wearability) and processing software (e.g. efficiency, robustness and intelligibility speech generation). Preliminary discussions on the influential factors affecting the SSIs' implementation were presented in [8], based upon criteria such as ability to operate in silence and noisy environments, usability by laryngectomees, issue of invasiveness market reediness and cost.

In the present work we employ the Permanent Magnet Articulography (PMA), which is a type SSI that is based on sensing the changes in the magnetic field generated by a set of permanent magnet markers attached onto the vocal apparatus (i.e. lips and tongue) during speech articulation by using an array of magnetic sensors located around the mouth [1, 3]. Although PMA shares some similarities with Electromagnetic Articulography (EMA) [5, 17], it does not explicitly provide the Cartesian position/ orientation of the markers, but rather a summation of the magnetic fields from magnets that are associated with a particular articulatory gesture. The focus here is to build upon our previous work of [20], to further improve and alleviate the shortcomings from a hardware perspective. The proposed prototype has several distinctive features, such as being miniature in size, highly portable, discreet and unobtrusive since it is hidden from sight within the user's mouth.

The rest of the chapter is organized as follows. Section 2 overviews the PMA technique and its development to date. Next, section 3 outlines the design challenges of the intraoral version of the PMA device. Then, section 4 describes the architecture of the intraoral PMA prototype. Section 5 describes the experimental methods used to assess performance, followed by the results of that evaluation in section 6. The final section concludes this chapter and provides an outlook for future work.

2 System Overview

PMA is a sensing technique for capturing the magnetic field resulting from movement of a set of permanent magnets attached onto the lips and tongue during speech articulation. The variations of the magnetic field can then be used to determine the speech which the user wishes to produce by first performing automatic speech recognition (ASR) on the PMA data and then synthesising the recognised text using a text-to-speech (TTS) synthesizer [3, 18-20].

A number of PMA prototypes have been investigated in recent years. Earlier prototypes [3, 18-19] provided acceptable speech recognition performance, but were not particularly satisfactory in terms of their appearances, comfort and ergonomic factors for the users. To address these challenges, a PMA prototype in the form of a wearable headset (designed based on a customized pair of spectacles or a headband) comprising of miniaturized sensing modules and wireless capability was developed [20]. The second generation prototype was re-designed based on a user-centered approach utilizing feedback from questionnaires completed by potential users and through discussion with stakeholders including clinicians, potentials users and their families. The appearance and comfort of the prototype was much improved and it demonstrated comparable recognition performances to its predecessors.

As illustrated in Fig. 1, the second generation PMA system consists of a set of six cylindrical Neodynium Iron Boron (NdFeB) permanent magnets, four on the lips (ø1 mm × 5 mm), one at the tongue tip (ø2 mm × 4 mm) and one on the tongue blade (ø5 mm × 1 mm). These magnets are currently attached using Histoacryl surgical tissue adhesive (Braun, Melsungen, Germany) during experimental trials, but will be surgically implanted for long term usage. The remainder of the PMA system is composed of a set of four tri-axial Anisotropic Magnetoresistive (AMR) magnetic sensors mounted on the wearable headset, a set of microcontrollers, rechargeable battery and a processing unit (e.g. computer/ tablet PC). Detailed information on these hardware modules and their operations is presented in [20].

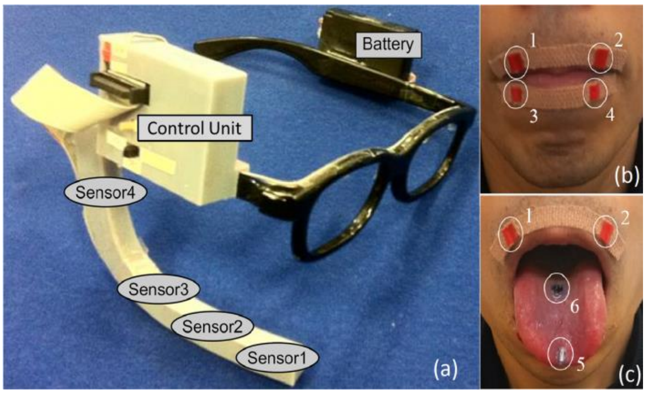


Fig. 1. (a) A wearable PMA prototype designed in a form of spectacles. (b) & (c) Placement of six magnets on lips (pellets 1-4), tongue tip (pellet 5) and tongue blade (pellet 6).

3 Design Challenges

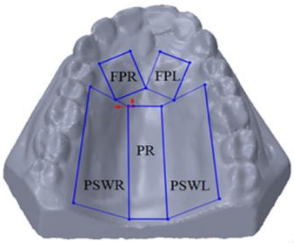
Although the second generation prototype has many desirable hardware features, it is not without drawbacks. Firstly, the performance of the external headset cannot be maintained in certain real-life conditions (i.e. exaggerated movement or sports activity) due to issues with instability. If there is a considerable movement of the headset on the user's head, the PMA system may need re-calibration/re-training to avoid degradation in performance. In addition, wearing the headset over long periods may not be comfortable, despite the fact that the device was designed to be lightweight and ergonomically friendly. Lastly, and potentially most importantly the external version of the PMA device may still be cosmetically unacceptable to some users. Previous studies indicated that the appearance is one of the most important factors that affect the acceptability of any AT by their potential end users [21-23].

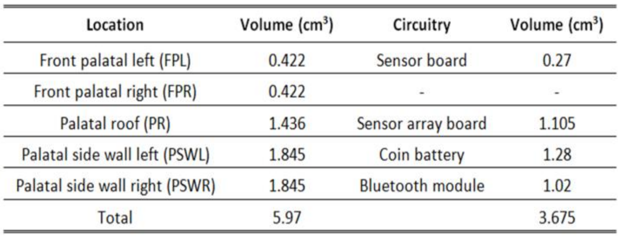
In order to overcome these limitations, an intraoral version of the PMA prototype, which fits under the palate inside the user's mouth in a form of a dental retainer, was proposed. Being tightly clamped onto the upper teeth means that the device would be more stable than the previous wearable headset. Due to the fact that the device is completely hidden from sight during normal use, it is cosmetically inconspicuous. In addition, since the sensors are much closer to the articulators than the external headset, the size of the implants can be significantly reduced. Similar intraoral-based designs have been previously implemented for other non-speech related ATs with various degree of success [24-26].

4 System Description

4.1 Space Budget

The latest intraoral-based PMA system is made up of: three tri-axial magnetic sensors, a wireless communication module, a microprocessor to synchronize data capture and communications and a suitable power source capable of providing an appropriate operating lifetime. This must be accommodated within the oral cavity, without excessively interfering with the natural tongue articulation during speech. A recent study [27] suggested that the palatal cavity is suitable to house the intraoral circuitry because of its relatively flat surfaces and proximity to the articulators. As illustrated in Fig. 2, a 3D palatal model was created and divided into five possible locations to accommodate the intraoral circuitry: front palatal left (FPL), front palatal right (FPR), palatal roof (PR), palatal side wall left (PSWL) and palatal side wall right (PSWR). The estimated space available in the palatal cavity on our test subject is 5.97cm^3 as shown in Fig. 2 (assuming a uniform 3mm thickness), whereas the estimated volume of the intraoral circuitry, as described subsequently, is approximately 3.68 cm^3.





Location	Volume (cm³)	Circuitry	Volume (cm³)
Front palatal left (FPL)	0.422	Sensor board	0.27
Front palatal right (FPR)	0.422	-	-
Palatal roof (PR)	1.436	Sensor array board	1.105
Palatal side wall left (PSWL)	1.845	Coin battery	1.28
Palatal side wall right (PSWR)	1.845	Bluetooth module	1.02
Total	5.97		3.675

Fig. 2. Space within the palatal cavity.

4.2 Intraoral Circuitry

A crucial design element for the intraoral circuitry is to drastically reducing the size of the electronics and rechargeable battery of the external version of PMA prototype, so that all necessary circuitry can be fitted inside the mouth. The major components of the PMA prototype are shown in Fig. 3. These are implemented using a low-power ATmega328P microcontroller, three tri-axial HMC5883L magnetic sensors (AMR), a rechargeable Li-Ion coin battery (capacity of 40mAh, 3.7V and 20 mm diameter × 3.2 mm thickness), and a wireless transceiver (Bluetooth 2.0 module). The remainder of the system shown in Fig. 4 consists of a processing unit (e.g. computer/ tablet PC) and a set six permanent magnets (NdFeB) attached onto lips and tongue in the same locations as illustrated in Fig. 1. The elements of the intraoral sensing system (which have a total volume of 3.68 cm^3) are arranged as shown in Fig. 4(a). These may be encapsulated and placed in the oral cavity as shown in Fig. 4(d).

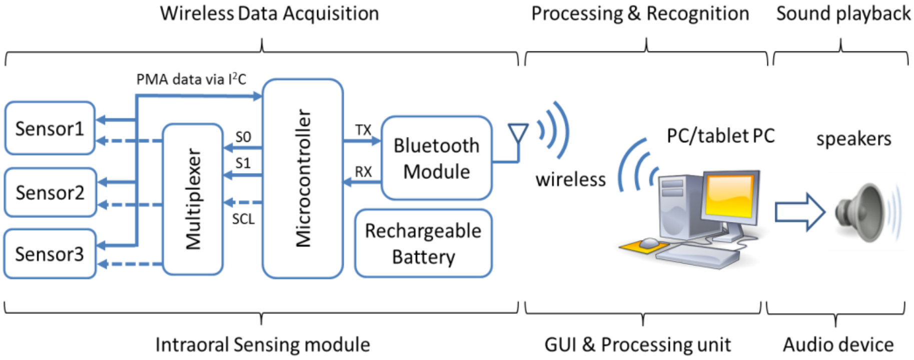


Fig. 3. Simplified operation block diagram.

Although there were many design changes for the intraoral design, the positions of the magnets remained unchanged from the earlier prototype. However, because of the proximity of the sensors, significantly smaller magnets (see Fig. 4c) can be used (note that the magnetic field strength decreases with cube of the distance away from the magnets): four on lips (ø1 mm × 4 mm), one on the tongue tip (ø1 mm × 1 mm) and one on the tongue blade (ø1 mm × 1 mm).

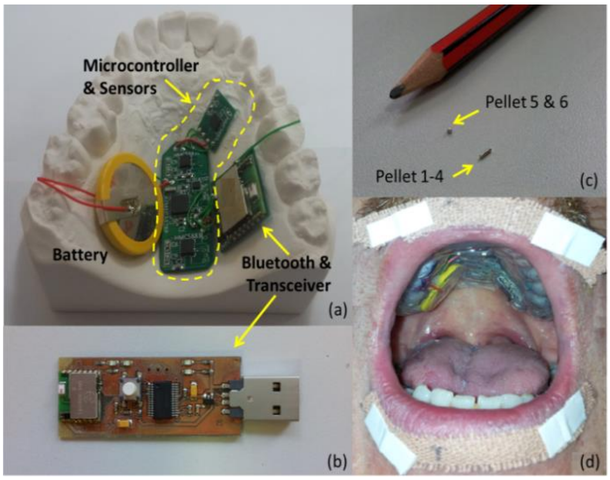


Fig. 4. (a) & (b) Circuitry of the intraoral version of the PMA system. (c) Placement of magnets on lips (pellets 1-4), tongue tip (pellet 5) and tongue blade (pellet 6). (d) View of the device when worn by the user.

4.3 Circuit Operation

The operational block diagram of the intraoral version of the PMA system is presented in Fig. 3 . A command is sent wirelessly from the processing unit to the intraoral sensing module via Bluetooth to trigger data acquisition. All three tri-axial magnetic sensors then measure the three components of magnetic field and digitize it with 12-bit resolution. The microcontroller acquires these measurements (9 PMA channels sampled at 80 Hz) through managing a multiplexer using three control signals (S0, S1 and SCL). The multiplexer acts as a switching device to route the serial clock (SCL) to the desired magnetic sensor through the I^2C interface. The acquired samples are then transmitted back to the processing unit wirelessly via the Bluetooth transceiver and custom designed Bluetooth dongle (in Fig. 3(b)) for further processing. Unlike the external version of the PMA prototype, the intraoral device is restricted to only operate wirelessly from inside the mouth. Hence, wired connectivity is impossible, as the sensing modules are to be sealed and packaged inside a dental retainer. In terms of software, a bespoke MATLAB-based graphical user interface (GUI) developed in [20] was adapted, where all speech processing and recognition algorithms were embedded.

4.4 Power Budget

As the circuitry is to be sealed into a dental retainer, the only way the intraoral device can acquire power is from a battery. With limited space available, only a small battery can be accommodated (in the current design, the battery takes 27% of the total volume of the circuitry). The battery can be recharged through a charging point located on the under-side of the dental retainer. In addition, any measures to extend the battery life will be of interest. Power hungry components such as the microcontroller, the magnetic sensors and the Bluetooth module may be set to *standby mode* or *sleep mode* to reduce the current consumption when they are inactive. As shown in Table 1, *sleep mode* gives a saving of 93% over *standby mode* or a saving of 97% over *active mode*.

Table 1. Current consumption in difference operational modes.

Current Consumption	Active mode (mA)	Stanby mode (mA)	Sleep mode (mA)
Sensors	5.1	0.006	0.006
Microcontroller	5.4	4.4	0.7
Bluetooth	19.0	7.22	0.007
Total	29.5	11.626	0.776

Fig. 5 shows a summary of the discharging cycle of the battery with the circuit in *active* and *sleep* modes. Neither of these operating regimes is fully representative of the expected use since they correspond to continuous speech and no speech respectively. If the system is to operate continuously (in *active mode*), the battery will last approximately one hour before being depleted below the minimum operating voltage (cut-off voltage) required by the Bluetooth module of 2.1V. In contrast, if the system was inactive at all times (in *sleep mode*) the battery would last about 32 hours. Based on the measurements in Table 1 and Fig. 5, a more realistic regime would be to allow 30 minutes of speech with a further 16 hours in *sleep mode*. Hence, the estimated usage time is considered to be sufficient for a typical day before charging is required. This assumes that the circuit is active only while utterance is underway, which implies that a user interface is required to allow speech to be initiated. Note that the intraoral circuit can be 'woken up' by Bluetooth command sent from the processing unit, so a variety of user interfaces could be devised.

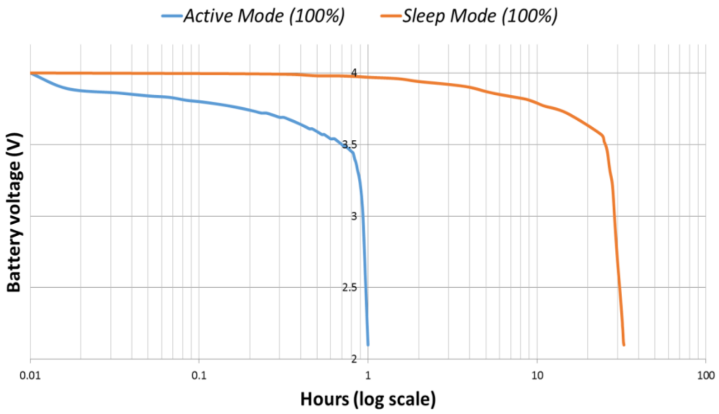


Fig. 5. Battery discharging over time under *active mode* and *sleep mode*.

4.5 System Implementation

The intraoral circuitry described above must be encapsulated to protect it from damage and short circuits due to saliva and to ensure it is held in place within the palate. The retainer must be customized according to the individual's oral anatomy. This may be achieved by forming it on a dental impression of the user's oral cavity (seen in the background of Fig. 4a). The intraoral PMA prototype was implemented in the form of dental retainers utilizing both soft and semi-rigid materials, as illustrated in Fig. 6. We will refer to these as *Type I* and *Type II*, respectively.

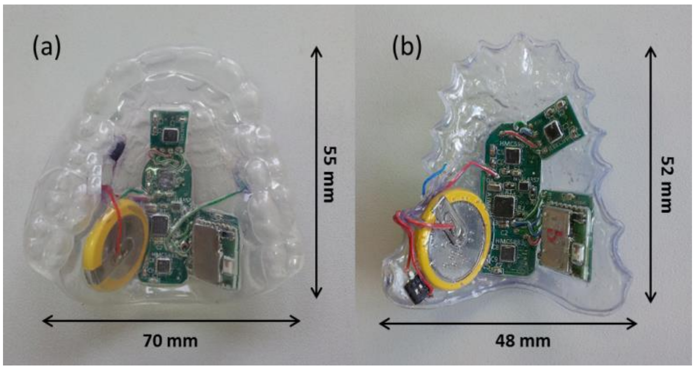


Fig. 6. PMA circuitry embedded inside a (a) soft bite raiser like dental retainer (*Type I*) and (b) semi-rigid dental retainer (*Type II*).

Type I (soft) retainer is similar to a soft bite raising appliance and is made of polypropylene or polyvinylchloride (PVC) material. On the other hand, *Type II* (semi-rigid) retainer is made from Essix C+ plastic. To allow stable fitting in the palate, the *Type I* retainer is fitted over the entire arch of the upper teeth. In contrast, the *Type II* retainer utilizes a set of curved edges to clasp tightly onto the upper teeth.

In generally, both intraoral and external PMA devices are speaker dependent systems, because their designs need to be individually tailored, based on the user's head or oral anatomy for optimal performance. In the case of the external device, this involves moving sensor arm so that it is close to the user's cheek and lips while in the case of the intraoral device, it must be encapsulated and formed on an impression of the user's palate.

5 Methods

5.1 Experimental Design

The data used for evaluating the new intraoral prototype were collected from a male native English speaker who is proficient in the usage of the external PMA device. Magnets were temporary attached on the subject using Histoacryl surgical tissue adhesive (Braun, Melsungen, Germany).

Recordings of PMA and audio data for training and evaluation were performed via using a customized Matlab GUI. The software provides a visual prompt of randomized utterances to the subject at interval of 5 seconds during the training session. The subject's head was not restrained during the recording sessions, but the subject was requested to avoid any large head movements. This was necessary to ensure that interference induced by movement relative to earth's magnetic field was at its minimum, so that it did not corrupt or distort the desired signal. This is because the current prototype is not yet equipped with a background cancellation/removal mechanism.

The recordings were conducted in an acoustically isolated room for optimal sound quality. The audio data were recorded using a shock-mounted AKG C1000S condenser microphone via a dedicated stereo USB-sound card (Lexicon Lambda) to a PC, with a 16 kHz sampling rate. Meanwhile, the PMA data were captured at a sampling frequency of 80 Hz via the intraoral PMA device and transmitted to the same PC wirelessly via Bluetooth, as illustrated in Fig. 3. Since both data streams (PMA & audio) are acquired from separate modality, synchronization between the two data streams is necessary. Therefore, an automatic timing re-alignment mechanism was implemented utilizing start-stop markers generated in additional to both data streams.

5.2 Data Corpus and Recording

Our long term goal is to explore the feasibility of using the intraoral device for continuous speech reconstruction. For preliminary testing, the TIDigits database [28] was selected because the limited size of the vocabulary enables whole-word model training from relatively sparse data and because of the simplicity of the language involved. The corpus consists of sequences of connected English digits with up to seven digits per utterance. The vocabulary is made up of eleven individual digits, i.e. from 'one' to 'nine', plus 'zero' and 'oh' (both representing digit 0).

The experimental data were collected from two independent sessions, with each session consisted of four datasets containing 77 sentences each. A total of 308 utterances containing 1012 individual digits were recorded during each session. To prevent subject fatigue, short breaks in between each recording session were allowed.

5.3 HMM Training and Recognition

Prior to the training and recognition processes, the acquired PMA data were segmented and checked using the audio data. Inappropriate endpoints were manually corrected if necessary. In addition, any mislabeled utterances were corrected using the acquired audio data.

The PMA data was then subjected to offset removal via median subtraction over 2s windows with 50% overlap and followed by data normalization. Next, the delta parameters were computed for all PMA channels and added to its original time series data, resulting in a feature vector of size 18. The delta-delta parameters were not included as part of the feature vector as they did not produced significant improvement in performance [18, 19]. The recognition performance based on the audio data was also evaluated for comparison purposes. In this case, 13 Mel-frequency cepstral coefficients (MFCCs) were extracted from the audio signals using 25ms analysis windows with 10ms overlap. Next, the delta and delta-delta parameters were computed and appended to the static parameters, resulting in a feature vector of dimension 39. An overview on the PMA and audio parameters used is presented in Table 2.

Table 2. Vector sizes of the parameters used in PMA and audio.

Parameters	Original	1st delta	2nd delta	Vector size
Sensor	×			9
SensorD	×	×		18
Audio	×	×	×	39

The extracted PMA and audio features were used for training two independent speech recognizers using the HTK toolkit [29]. In both cases, the acoustic model in the recognizer uses whole-word Hidden Markov Models (HMMs) [30] for each of the eleven digits. Each HMM has 21 states and 5 Gaussians per state. The selected parameters were not optimized, but were known for their performances based on previous work [18, 19]. The HMM training and recognition was carried out in four validation cycles. In each cycle, three out of four sets within a session were used for training and the remaining one for testing. The recognition results were averaged over four cycles and across two independent sessions.

6 Results and Discussion

6.1 Evaluation of the Intraoral Devices

As seen in Fig. 7, it is obvious that *SensorD* performs significantly better than using *Sensor* data alone across both *Type I* and *Type II* intraoral devices. Similar trends where *SensorD* is superior over *Sensor* were also reported in [7]. In addition, the *Type II* intraoral design outperformed its counterpart (i.e. *Type I*) on both word and sequence recognition. Although the hardware on both devices were similar, the *Type II* device had its front sensor placed to the side, whereas it was positioned at the center for the *Type I* design. This is to eliminate or at least minimize possible saturation at the front sensor due to contact with magnet attached onto the tongue tip. Since the *Type II* intraoral version provided superior performance, this version will be the focus for the rest of this chapter.

Fig. 8 illustrates that an increased number of training sessions yields better performance on both word and sequence recognitions, through the reduction of in word error rate (WER). 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.

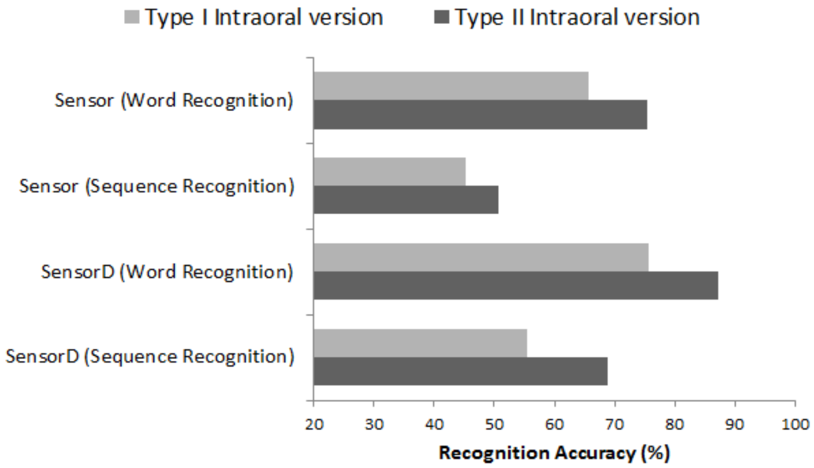


Fig. 7. Comparison of word and sequence accuracies of connected digits between *Type I* and *Type II* intraoral version.

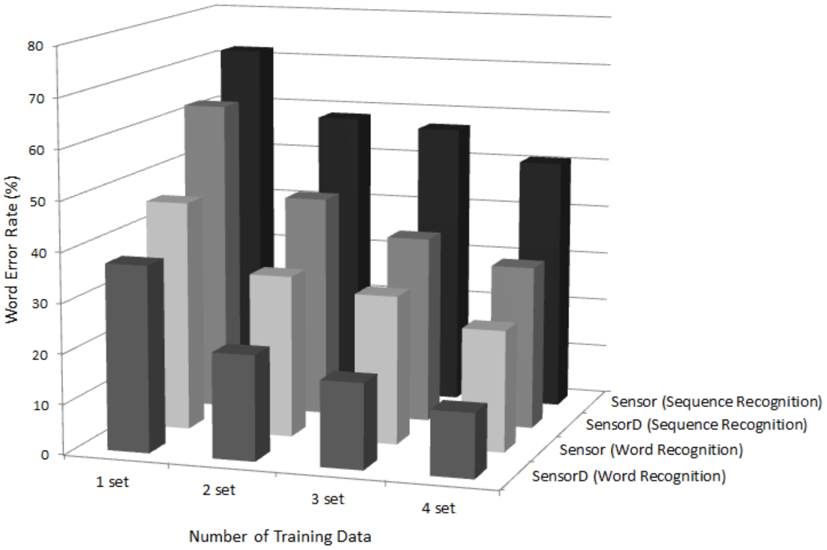


Fig. 8. Decrease in word error rate (WER) with the increase in training sessions.

6.2 Recognition Performance

Both word and sequence recognition results for the intraoral and external versions of the PMA device are presented in Fig. 9 and Fig. 10. In addition, the performances of the PMA devices were compared with audio-based recognition. The darker bars indicate the performance achieved using only static PMA data (vector size of 9), whereas the lighter bars are the results achieved using both static and dynamic features (vector size of 18). In addition, the grey-colored bars are the speech-recognition performance achieved using audio data (vector size of 39). We will refer to these three conditions as *Sensor*, *SensorD* and *Audio* features, respectively (see Table 2).

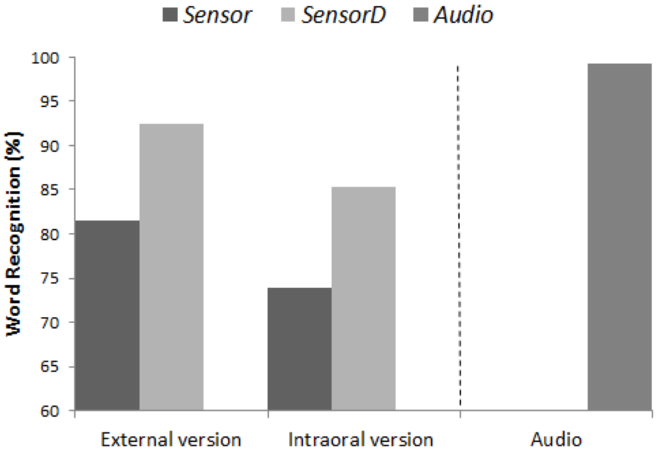


Fig. 9. Comparison of word accuracy in the connected digits.

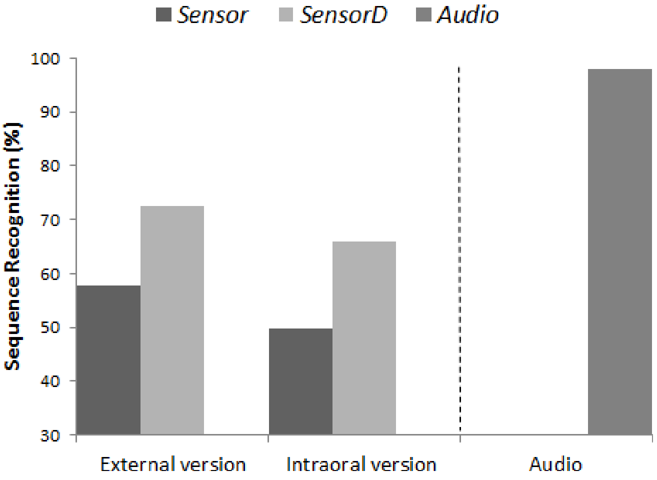


Fig. 10. Comparison of sequence accuracy in the connected digits.

The results reflect the mean of the data collected across the two sessions, but were initially analyzed independently session-by-session. In order to avoid the inconsistency of magnets placement during individual training sessions, data were not merged across different sessions. This however could be solved, as magnets are to be surgically implanted for long term usage. Alternatively, session-independent approaches such as those presented for other SSIs methods could be investigated [14, 31].

As shown in both Fig. 9 and Fig. 10, it is quite obvious that *SensorD* produced better recognition performance on both occasions than using *Sensor* alone Similar trends were also reported [18, 20]. As expected, for this simple task, recognition using *Audio* performed very well (i.e. 99%). Preliminary evaluations indicate a close comparable recognition performance for the intraoral device and the previous external version, as illustrated in Fig. 9 and Fig. 10. There are a number of possible explanations for this degradation: 1) the presence of the intraoral prototype affects articulation and, in particular limits the tongue movements. This may lead to inconsistent articulation, 2) the subject was new to the intraoral version, but had prior experiences on the external PMA version, 3) possible drawbacks of operating at a lower sampling rate (up to 80Hz) due to the design constraint on the intraoral device. Although recognition performance decreases with the used of lower sampling rate, both external and intraoral version showed similar recognition trends (illustrated in Fig. 11), and 4) the magnets are able to come much closer to the sensors in the intraoral device than in the external device, resulting in a more significant non-linear effect (since the field strength decreases with cube of the distance). This means that small unintentional articulator movements (e.g. swallowing, licking the lips and etc.) can generate very large signals in some instances which could have corrupted the data. Further work is required to understand the significance of each of these possible causes.

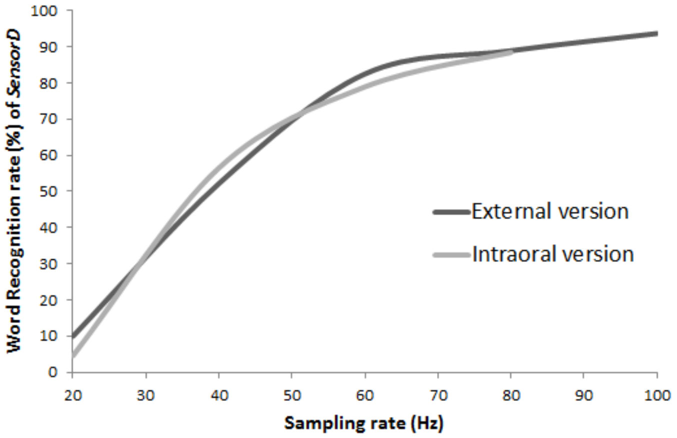


Fig. 11. Decrease in recognition performance with the reduction of sampling rate.

6.3 Hardware Comparison

As discussed in section 3, one major obstacle to the acceptability of an AT (e.g. SSI) is its appearance if it is considered unattractive. Similar views were also concluded through discussions with potential users who have undergone a laryngectomy and an opinion survey of 50 laryngectomees and their families/ friends: the appearance of the device was considered to be of a very high priority [20]. To enhance its appeal to users, influential factors such as appearance need to be accounted for during device development. The challenge here is to satisfy the design objective and continue improving the PMA device's appearance but without compromising its speech reconstruction performance. The latest intraoral prototype employs the same functional principles as the previous design reported in [20, 32], but implemented in a different form. A summary of the hardware features of the new intraoral PMA system compared to its predecessor is presented in Table 3.

Table 3. Summary of the PMA devices' specifications and comparison [*Note that although the external sensing system has 12 channels, only 9 are used for speech recognition and 3 are used for cancellation of background magnetic fields]**.**

Specifications		**Intraoral Device**	**External Device**
Appearance		Dental retainer	Wearable headset
Operating voltage		2.1 V	5 V
Magnets	Tongue Blade	ø1 mm × 1 mm	ø5 mm × 1 mm
	Tongue Tip	ø1 mm × 1 mm	ø2 mm × 4 mm
	Lips	ø1 mm × 4 mm	ø1 mm × 5 mm
Magnetic Sensing	Dimension	12 × 12 × 3 mm^3	12 × 12 × 3 mm^3
	Sensitivity	230 LSb/gauss	440 LSb/gauss
	Sampling rate	80 Hz	100 Hz
	Channels	9	12*
Data Transmission	Type	Bluetooth 2.0	Bluetooth 2.0/ USB
	Frequency	2.4 GHz	2.4 GHz
	Data rate	57.6 kbps	500 kbps
Power	Supply	Rechargeable battery	Rechargeable battery/ USB
	Battery	Li-Ion 40 mAh	Li-Ion 1080 mAh
	Current consumption	30.5 mA	93.5 (wireless) / 67.1 (wired) mA
	Lifetime	1 hour	10 hours
Prototype	Dimension	70 × 55 × 25 mm^3	160 × 160 × 150 mm^3
	Weight	15 g	160 g
	Material	Polypropylene / Essix C+ plastic	VeroBlue / VeroWhitePlus resin

Despite the improved appearance of the second generation PMA system in the form of a wearable headset, it might not yet to be appealing to all potential users. To address this shortcoming, the latest intraoral circuitry was implemented in the form of a dental retainer. To achieve this, the circuit was re-designed to use fewer and smaller components. In addition, the power consumption of the circuit was carefully managed to allow it to operate from a small battery suitable for inclusion within the dental retainer. Hence, this led to a much smaller and lighter (i.e. one tenth of previous weight) prototype as compared to its predecessor. In addition, the intraoral prototype is highly portable, it operates and can be controlled wirelessly via Bluetooth using a computer/tablet PC. Also, a higher signal-to-noise ratio (SNR) was obtained with smaller magnetic markers, due to their proximity to the magnetic sensors. The tongue magnets used with the intraoral sensor system had 16 to 25 times smaller volume than those used for the external headset, potentially making them less invasive when implanted.

A significant drawback with the intraoral device is the limited battery size and capacity (i.e. 40mAh). In contrast, the external version is less restricted in term of size and weight of the battery. Hence, this significantly reduces the operational time of the intraoral device per charging. A number of steps have been introduced to reduce its power consumptions: a lower operating voltage is selected and power-efficient components, lower data sampling and transmission rates were chosen. In addition, software was developed to switch from an *active mode* to *sleep mode* when not in use. Using these measures, it is estimated that the battery life cycle could be extended from one hour to about 16.5 hours including 30 minutes of speech.

7 Conclusion

In this chapter we have described a new intraoral PMA prototype using commercial off-the-shelf (COTS) components embedded inside a dental retainer constructed using the subject’s dental impression. Preliminary evaluation of the intraoral prototype indicated a near comparable recognition performance to previous external sensor systems.

Although the intraoral version showed minor degradation in performance, there are several advantages over its predecessor and with a number of avenues for further investigation to improve its performance. It is also considered to be more stable and robust against unintentional movement as it is implemented in a form of a dental retainer, which securely sits in the palatal cavity and is clasping firmly on the upper teeth. Secondly, significantly smaller magnets may be used for the intraoral version (because of their proximity to the magnetic sensors) while also giving a higher SNR. In addition, the dental retainer can be completely hidden inside the user’s mouth and out of sight. Hence, this would eliminate the concern of being a sign of disability. However, a downside of the intraoral design would be the possibility of limiting the natural movement of the tongue, because the device occupies part of the user’s oral cavity. Further work is required to assess whether users become accustomed to the presence of the device and are able to achieve more consistent articulation.

With these encouraging results obtained, extensive work is needed to: 1) further reduce the size of future intraoral prototypes, 2) improve the circuitry power efficiency, 3) incorporate inductive charging for the battery, and 4) introduce a background cancellation mechanism for movement-induced interference. Though there are still limitations, the present work demonstrates a major step towards creating a viable SSI that would appeal to speech impaired users.

Lastly, an alternative speech generation through direct conversion of PMA data into audible speech without an intermediate recognition step was investigated and preliminary results were encouraging [33]. For further information on the PMA-based SSI and its speech restoration technique, please visit www.hull.ac.uk/speech/disarm.

Acknowledgements. The authors would like to thank Helen Dehkordy from Hull and East Yorkshire Hospitals NHS Trust for prototyping the dental retainers. The work is an independent research funded by the National Institute for Health Research (NIHR)'s Invention for Innovation Programme (Grant Reference Number II-LB-0814-20007). 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. Braz, D. S. A., Ribas, M. M., Dedivitis, R. A., Nishimoto, I. N., and Barros, A. P. B.: Quality of life and depression in patients undergoing total and partial laryngectomy. Clinics, 60(2), 135-142 (2005).
3. 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).
4. Liu, H., and Ng, M.: Electrolarynx in voice rehabilitation. Auris Nasus Larynx, 30(3), 327-332 (2007).
5. Wang, J., Samal, A., Green, J. R., and Rudzicz, F.: Sentence recognition from articulatory movements for silent speech interfaces. In: Proceedings of 37th ICASSP, pp. 4985-4988, Kyoto, Japan (2012).
6. Toda, T., Nakagiri, M., and Shikano, K.: Statistical voice conversion techniques for body-conducted unvoiced speech enhancement. IEEE Transactions on Audio, Speech and Language Processing, 20(9), 2505-2517 (2012).
7. Doi, H., Nakamura, K., Toda, T., Saruwatari, H., and Shikano, K.: Esophageal speech enhancement based on statistical voice conversion with Gaussian mixture model. IEICE Transactions on Information and Systems, 93(9):2472-2482 (2010).
8. 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).
9. Brumberg, J.S.; Wright, E.J.; Andreasen, D.S.; Guenther, F.H.; Kennedy, P.R.: Classification of intended phoneme production from chronic intracortical microelectrode recordings in speech-motor cortex. Frontiers in neuroscience, 65(5), 1–12 (2011).
10. Brumberg, J. S., Nieto-Castanon, A., Kennedy, P.R., and Guenther, F. H.: Brain-computer interfaces for speech communication. Speech Communication, 52(4), 367-379 (2010).
11. Porbadnigk, A., Wester, M., Calliess, J., and Schultz, T.: EEG-based speech recognition–impact of temporal effects. In: Proceedings of 2nd BIOSIGNALS, pp. 376-381, Porto, Portugal (2009).
12. Jou, S. C. S., Schultz, T., Walliczek, M., Kraft, F., and Waibel, A.: Towards continuous speech recognition using surface electromyography. In: Proceedings of 9th INTERSPEECH, pp. 573-576, Pittsburgh, USA (2006).
13. Wand, M., Janke, M., and Schultz, T: Tackling speaking mode varieties in EMG-based speech recognition. IEEE Transaction on Biomedical Engineering, 61(10), 2515-2526 (2014).
14. Wand, M., and Schultz, T.: Session-independent EMG-based speech recognition. In: Proceedings of 4th BIOSIGNALS, pp. 295-300, Rome, Italy (2011).

15. Petajan, E. D.: An architecture for automatic lipreading to enhance speech recognition. In: Proceedings of the International Conference on Computer Vision and Pattern Recognition, pp. 40-47, California, USA (1985).
16. Hueber, T., Benaroya, E.-L., Chollet, G., Denby, B., Dreyfus, G., and Stone, M.: Development of a silent speech interface driven by ultrasound and optical images of the tongue and lips. Speech Communication, 52(4), 288-300 (2010).
17. Toda, T., Black, A. W., and Tokuda, K.: Statistical mapping between articulatory movements and acoustic spectrum using a Gaussian mixture model. Speech Communication, 50(3), 215-227 (2008).
18. 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).
19. 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 14th INTERSPEECH, pp. 1140-1143, Lyon, France (2013).
20. Cheah, L. A., Bai, J., Gonzalez, J. A., Ell, S. R., Gilbert, J. M., Moore, R. K., and Green, P. D.: A user-centric design of permanent magnetic articulography based assistive speech technology. In: Proceedings of 8th BIOSIGNALS, pp.109-116, Lisbon, Portugal (2015).
21. 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).
22. 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), 49-62 (2006).
23. Bright, A. K., and Conventry, L.: Assistive technology for older adults: psychological and socio-emotional design requirements. In: Proceedings of 6th International Conference on PErvaesive Technologies Related to Assistive Environments, pp. 1-4, Rhodes, Greece (2013).
24. Tang, H., and Beebe, D. J.: An oral interface for blind navigation. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 14(1), 116-123 (2006).
25. Lontis, E. R., Lund, M. E., Christensen, H. V., Gaihede, M., Caltenco, H. A., and Andreasen-Strujik, L. N.: Clinical evaluation of wireless inductive tongue computer interface for control of computers and assistive devices. In: Proceedings of 32nd IEEE EMBC, pp. 3365-3368, Beunos Aires, Argentina (2010).
26. Park, H., Kiani, M., Lee, H. M., Kim, J., Block, J., Gosselin, B., and Ghovanloo, M.: A wireless magnetoresistive sensing system for an intraoral tongue-computer interface. IEEE Transactions on Biomedical Circuits and Systems, 6(6), 571:585 (2012).
27. Bai, J., Cheah, L. A., Ell, S. R., and Gilbert, J. M.: Design of an intraoral device based on permanent magnetic articulography. In: Proceedings of Macau Conference on Engineering, Technology and Applied Science, pp. 1-12, Macau, China (2015).
28. Leonard, R. G.: A database for speaker-independent digit recognition. In: Proceedings of 9th ICASSP, pp. 328-331, San Diego, USA (1984).
29. 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: Cambridge University Press (2009).
30. Rabiner, L.R.: A tutorial on Hidden Markov Models and selected applications in speech recognition. Proceedings of the IEEE, 77, 257-286 (1989).

31. Maier-Hein, L., Metze, F., Schultz, T., and Waibel, A.: Session independent non-audible speech recognition using surface electromyography. In: Proceedings of Automatic Speech Recognition and Understanding Workshop, pp. 331-336, Cancun, Mexico (2005).
32. 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 15th INTERSPEECH, pp. 1018-1022, Singapore (2014).
33. Gonzalez, J. A., Cheah, L. A., Gilbert, J. M., Bai, J., Ell, S. R., Green, P. D., and Moore, R. K.: Direct speech generation for a silent speech interface based on permanent magnet articulography. In: Proceedings of 9th BIOSIGNALS, pp. 109-116, Lisbon, Portugal (2016).