

# Permanent Magnetic Articulograph (PMA) vs Electromagnetic Articulograph (EMA) in Articulation-to-Speech Synthesis for Silent Speech Interface

Beiming Cao<sup>1</sup>, Nordine Sebki<sup>3</sup>, Ted Mau<sup>4</sup>, Omer T. Inan<sup>3</sup>, Jun Wang<sup>1,2</sup>

<sup>1</sup>Speech Disorders & Technology Lab, Department of Bioengineering

<sup>2</sup>Callier Center for Communication Disorders

University of Texas at Dallas, Richardson, TX, USA

<sup>3</sup>Inan Research Lab, School of Electrical and Computer Engineering  
Georgia Institute of Technology, Atlanta, GA, USA

<sup>4</sup>Department of Otolaryngology - Head and Neck Surgery  
University of Texas Southwestern Medical Center, Dallas, TX, USA

## Abstract

Silent speech interfaces (SSIs) are devices that enable speech communication when audible speech is unavailable. Articulation-to-speech (ATS) synthesis is a software design in SSI that directly converts articulatory movement information into audible speech signals. Permanent magnetic articulograph (PMA) is a wireless articulator motion tracking technology that is similar to commercial, wired Electromagnetic Articulograph (EMA). PMA has shown great potential for practical SSI applications, because it is wireless. The ATS performance of PMA, however, is unknown when comparing with current EMA. In this study, we compared the performance of ATS using a PMA we recently developed and a commercially available EMA (NDI Wave system). Datasets with same stimuli and size that were collected from tongue tip were used in the comparison. The experimental results indicated the performance of PMA was close to, although not as equally good as that of EMA. Furthermore, in PMA, converting the raw magnetic signals to positional signals did not significantly affect the performance of ATS, which support the future direction in PMA-based ATS can be focused on the use of positional signals to maximize the benefit of spatial analysis.

## 1 Introduction

People who had a laryngectomy have their larynx surgically removed in the treatment of a condition such as laryngeal cancer (Bailey et al., 2006). The removal of the larynx, as a treatment of cancer, prevents laryngectomees from producing speech sounds and inhibit their ability to communicate. Current approaches for improving their ability to communicate include (intra- or extra-oral) artificial larynx (Baraff, 1994), tra-

cheoesophageal puncture (TEP) (Robbins et al., 1984), and esophageal speech (Hyman, 1955). All of these approaches generate abnormal speech like hoarse voicing by tracheoesophageal speech or robotic voicing by artificial larynx (Mau, 2010; Mau et al., 2012). These patients may feel depressed because of their health status and anxiety during social interactions, as they think that other people perceive them as abnormal, or they directly experience symbolic violence (Mertl et al., 2018). As a result, the development of communication aids that can produce normal-sounding speech is essential to improving the quality of life for patients in this population.

Silent speech interfaces (SSI) are devices which convert non-audio biological signals, such as movement of articulators, to audible speech (Denby et al., 2010). Unlike existing methods, SSIs are able to produce natural sounding synthesized speech and even have the potential to recover the patients' own voices. There are currently two types of software designs in SSI. One is a "recognition-and-synthesis" approach, which is to convert articulatory movement to text, and then drive speech output using a text-to-speech synthesizer (Kim et al., 2017). The other design is direct articulation-to-speech (ATS) synthesis, which is more promising for SSI application, because ATS can be real-time. Currently, the prominent methods for capturing articulatory motion data include: electromagnetic articulograph (EMA) (Schönle et al., 1987; Cao et al., 2018; Bocquelet et al., 2016), permanent magnet articulograph (PMA) (Gonzalez et al., 2014; Kim et al., 2018), ultrasound image (Csapó et al., 2017), surface electromyography (sEMG) (Diener et al., 2018), non-audible murmur (NAM) (Nakajima et al., 2003). All of these technologies have

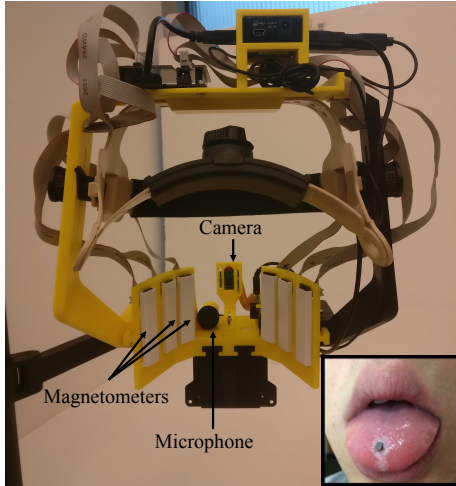


Figure 1: Our recently developed, head-set PMA device, where a small magnet is attached on the tongue tip.

their own advantages and disadvantages. PMA has recently shown its potential for SSI because it is wireless and suitable for future practical applications.

Unlike EMA that uses wired sensors attached on the articulators with a magnetic field generator outside, PMA attaches (wireless) permanent magnets to articulators and adopts magnetometers to capture the changes in the magnetic field generated by the motion of the magnets. These magnetic readings are then fed into a localization algorithm that estimates the 3D position of the magnet in the oral cavity (Sebkhi et al., 2017). Both EMA and PMA have been used in prior research on ATS (Cao et al., 2018; Gonzalez et al., 2017a; Cheah et al., 2018) with varying results. Although EMA has been shown to yield more precise measurements (Yunusova et al., 2009; Berry, 2011) compared to PMA (Sebkhi et al., 2017), EMA devices are normally cumbersome as they require wired sensors be attached to articulators. Additionally, EMA devices are normally expensive. In contrast, PMA devices are mostly very light and portable, relying on wireless tracking by using permanent magnets as the tracers, also affordable compared to EMA. Due to the wireless, portability and low-cost advantages of PMA, it offers an appealing alternative to EMA if it is able to achieve similar levels of performance as EMA in ATS systems. To our knowledge, however, no prior studies have directly compared the performance of these two technologies for SSI applications.

In this study, we compared the ATS perfor-

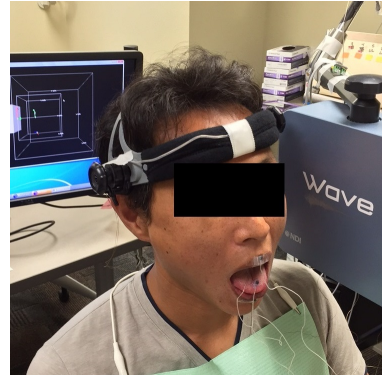


Figure 2: Wave System (EMA), where multiple sensors are attached on the tongue and lips. Only the tongue tip sensor data was used in the comparison with PMA.

mance of our recently developed PMA-based wireless tongue tracking system and a commercial EMA (NDI Wave system). We first examined whether it is more effective to use raw magnetic field signals than to use the converted magnet positional data ( $x, y, z$  coordinates) of PMA in ATS. Second, we compared the performance of EMA and PMA using tongue tip data only. A deep neural network (DNN)-based ATS model was used to evaluate the ATS performance for both EMA and PMA data. In this study, a dataset was collected from two groups of subjects who spoke the same stimuli using PMA or EMA, respectively. Tongue tip is the common flesh point in the PMA and EMA datasets, which were used for analysis in this study.

## 2 Dataset

### 2.1 PMA Data Collection

Ten subjects (6 males and 4 females, average age: 24.1 years  $\pm$  4.84) participated in the PMA data collection session in which they repeated a list of 132 phrases twice in their habitual speaking rate. The first repetition is normal voiced speech, and the second repetition is unvoiced speech. In this study, only the voiced speech data was used. The phrases in the list were phrases that are frequently spoken by users of augmentative and alternative communication (AAC) devices (Glennen and De-Coste, 1997). The PMA data was collected at the Georgia Institute of Technology.

The PMA data used in this study was collected with our newly developed wearable, headset system, which is based on the same magnetic technology in the prior benchtop version multimodal speech capture system (MSCS) (Sebkhi et al.,

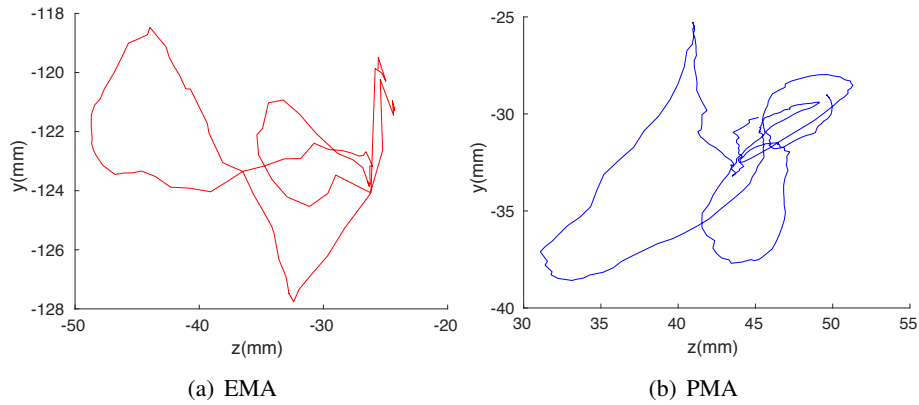


Figure 3: Lateral view samples of tongue tip trajectory captured by PMA and EMA when saying: “That is perfect!” (By two different subjects).

2017). Figure 1 shows the wearable, wireless tongue tracking system, which uses PMA and a camera for tongue and lip motion caption, respectively. A microphone was used for audio recording. This PMA system has an embedded array of magnetometers that measure the change of magnetic field generated by a magnetic tracer attached close to the tongue tip.

During a data collection session, a disk-shaped magnetic tracer (diameter = 3mm, thickness = 1.5mm, D21BN52, K&J Magnetics) was attached to about 1cm from tongue tip. An array of 24 external 3-axial magnetometers (LSM303D, STMicroelectronics) are divided into six modules, each with 4 magnetometers, which are positioned near the mouth, so there are two groups of 12 sensors that are near the right cheek and left cheek. These sensors were used for capturing the magnetic field fluctuations generated by the tracer, which are fed into a localization algorithm that estimates the 3D position of the magnet every 10 ms (100 Hz). The spatial tracking accuracy of the PMA varies from 0.44 to 2.94 mm depending upon the position and orientation of the tracer (Sebkhi et al., 2017). The audio data recording was sampled at 96000 Hz.

Previous studies (Gonzalez et al., 2017a; Cheah et al., 2018) show that the combination of multiple tracers on the tongue had better performance than single tracer (i.e., tongue tip). However, a smaller number of magnetic tracers on the tongue is critical for its practical use in daily life (Kim et al., 2018). Future users of this technology likely prefer to have only one permanent or semi-permanent attached magnetic tracer on their tongue. Even for lab experiment, attaching multiple tracers on the tongue takes longer time and relative logistic diffi-

culty to operate. In addition, with only one tracer on the tongue tip, the risk of accidentally biting it is very small (Laumann et al., 2015).

To provide the best tracking performance with one single tracer, the system relies on 24 magnetometers positioned outside the mount to accurately track the tongue motion (Kim et al., 2018). The six magnetometer modules are connected via serial peripheral interface (SPI) to a sensor controller module (Kim et al., 2018) that also includes a USB interface to communicate with the PC. More technical details about the tracking technology can be found in (Sebkhi et al., 2017). In this study, although wearable, the headset was anchored to a support in order to provide the best positional accuracy (to avoid possible head motion during recording).

## 2.2 EMA Data Collection

Another group of 10 gender- and age-matched subjects (6 males and 4 females, average age: 24.3 years  $\pm$  3.50) participated in the EMA data collection session. These individuals read the same list of 132 phrases used in the PMA data collection session. The EMA dataset was collected at the University of Texas at Dallas.

Wave system (Northern Digital Inc., Waterloo, Canada) was used for EMA data collection (Figure 2). Four small wired sensors were attached to the tongue tip (0.5 to 1cm from tongue apex), tongue back (20-30mm back from TT), upper lip and lower lip using dental glue or tape. Additionally, a fifth (head) sensor was attached to the middle of forehead for head correction. Finally, 3D EMA data was sampled at 100 Hz which is same to PMA data. The spatial precision of motion tracking is

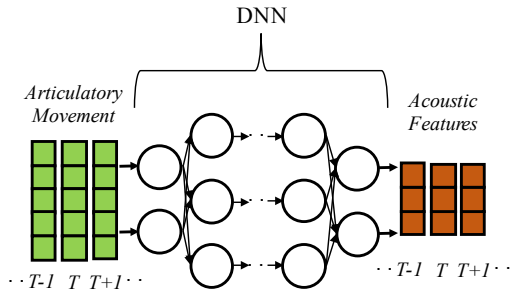


Figure 4: ATS using DNN.

about 0.5 mm (Berry, 2011), Figure 3(a) gives an example of two-dimensional (2D) EMA tongue tip movement trajectory (lateral view) when saying: “That is perfect!”. The sampling rate of audio data was 22050 Hz. NDI Wave system does not provide the raw magnetic signals.

To ensure an analogous comparison with the PMA device, only the tongue tip data collected using EMA was used in this study.

### 2.3 Data Preprocessing

To provide EMA and PMA consistent acoustic features, the sampling rates of audio data in EMA and PMA were resampled to same level. The audio data in PMA dataset was downsampled to 48000 Hz from 96000 Hz, and the audio data in EMA dataset was upsampled to 48000 Hz from 22050 Hz. After that, spectral envelope was extracted with Cheaptrick algorithm (Morise, 2015) and then converted to 60-dimensional mel-cepstral coefficients (MCCs) as the output acoustic features of ATS model. The MCCs were extracted at a rate of 200 frames per second, therefore, the PMA and EMA data were upsampled to 200 Hz to match the acoustic features.

Our PMA device captures the motion of tongue tip with the 72-channel raw magnet signals (3 axes 24 magnetometers). In addition to raw magnet signals, the 3D cartesian positions of the magnet tracer were obtained by localizing the raw magnet signals with nonlinear optimization method (Sebkhi et al., 2017). Figure 3(b) gives an example of a 2D trajectory (lateral view) of magnet tracer when saying “That is perfect!” obtained by localizing raw magnet signals. Both raw magnet signals and 3D-position signals were used in this study.

## 3 Method

### 3.1 Articulation-to-Speech Synthesis (ATS) Using Deep Neural Network (DNN)

The ATS model in this study uses a DNN to map articulatory signals (PMA or EMA) to acoustic features (MCCs) (Figure 4). The first and second order derivatives of both input articulatory and the output acoustic data frames were computed and concatenated to the original frames for context information.

The DNN has 6 hidden layers, each layer has 512 nodes with rectified linear unit (ReLU) activation function. During the DNN training, Adam optimizer (Kingma and Ba, 2014) was used, the maximum number of training epochs is 50, learning rate for PMA data is 0.008 and 0.005 for EMA data. The performances of ATS system is assessed using EMA positional data, PMA raw data, PMA positional data, and the combination of PMA raw and positional data. Therefore, the input dimensions of ATS in this study are: 9 (3-dim. PMA or EMA positional +  $\Delta$  +  $\Delta\Delta$ ), 216 (72-dim. PMA raw magnet signals +  $\Delta$  +  $\Delta\Delta$ ), and 225 (concatenation of 9-dim. and 216-dim.). The output dimension is 180 (60-dim. MCCs +  $\Delta$  +  $\Delta\Delta$ ). The DNN model in this study was implemented with Tensorflow machine learning library (Abadi et al., 2016).

### 3.2 Experimental Setup

As mentioned previously, we first compared the ATS performance using raw PMA signals, converted positional data, or both. This experiment will help to understand the which type of PMA data leads to the best performance. NDI Wave is a commercial system, which does not provide any magnetic signals that have not been localized, thus this experiment was conducted for our PMA system only. Second, we compared the best performance in PMA with the performance in EMA. The results will reveal which technology (PMA or EMA) performs better.

Speaker-dependent setup was used in both experiments, as speaker-independent ATS is considered challenging at this moment, due to the physiological difference among different speakers. The ATS performances on each subject were averaged as the final performance. For the 132 phrases in both PMA and EMA data, 110 phrases were used for training, 10 for validating, and 12 for testing.

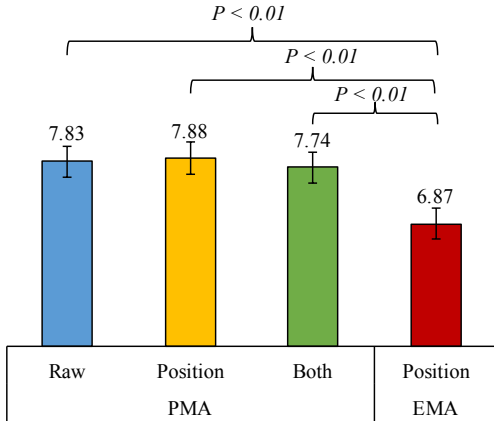


Figure 5: Average MCD of 10 PMA Subjects and 10 EMA Subjects. Statistical significances between the results using EMA and all types of PMA data on ATS model are computed with ANOVA tests.

The ATS results were measured with mel-cepstral distortion (MCD). MCD is calculated by equation (1), where  $C$  and  $C^{gen}$  denote the original and generated mel-cepstral coefficients (MCCs), respectively,  $m$  is the frame step (or time),  $d$  denotes  $d$ th dimension in frame  $m$ .  $D$  is the dimension of MCCs, which is 60 in this study.

$$MCD = \frac{10}{\ln 10} \sum_{m=1}^T \sqrt{2 \sum_{d=1}^D (C_{m,d} - C_{m,d}^{gen})^2} \quad (1)$$

As mentioned, lip movement information has not been used in this study, since PMA and EMA devices use different approaches for lip motion caption. PMA uses a computer vision algorithm to recognize the shape of the lips from images captured by an embedded camera, whereas EMA relies on tracking the motion of attached sensors to the vermilion borders of the lips to estimate lips gesture. In addition, due to the relatively small data size, the synthesized audio samples did not have sufficiently high speech intelligibility for listening test. Therefore, the subjective/listening testing was not conducted in this study.

## 4 Results and Discussion

### 4.1 Magnetic signals vs positional data in PMA

Experimental results are presented in Figure 5, where three-way ANOVA tests were used in the statistical analysis. First, for PMA, that performance using raw magnet data was not significantly

different to the performance using positional data only ( $p < 0.85$ ), and was also not significantly different with that using combined raw magnetic field signals and positional data ( $p < 0.76$ ). There was also no significance between the ATS performance using positional data only and that using combined raw magnetic field signals and positional data ( $p < 0.60$ ).

These findings suggest, for PMA, we could use either raw magnetic field signals or converted positional data for a similar level of performance. Combining these two signals together may not improve the performance. This finding is inconsistent with our prior study in silent speech recognition (SSR) using PMA data, where using magnetic signals outperformed than that using converted positional data (Kim et al., 2018). Further studies are needed to reveal why magnetic signals outperformed positional data in SSR, but their performance in ATS was not significantly different.

The finding that positional data can have similar performance with that using magnetic data is encouraging for our future development of ATS using PMA. Although mapping the raw magnetic signals directly to acoustic features is more straightforward, transforming these signals to positional signals allows the use of articulation data processing methods, such as Procrustes matching (Gower, 1975; Kim et al., 2017), that cannot be easily applied to the raw data. In addition, a PMA positional data-based ATS can be decoupled from a device configuration, it will be easier to change the number of sensors, their positions, their model, and their settings. Finally, a PMA positional data-based ATS has a potential of using EMA data for training, since they both track the 3D motion of articulators.

### 4.2 PMA vs EMA

Second, when comparing the ATS performance using PMA data and EMA data, the results obtained using PMA is not as equally good as that obtained in EMA. The performance in EMA significantly outperformed all the three configurations in PMA (raw, positional, and raw + positional data) ( $p < 0.01$  also in an ANOVA test).

Although the EMA-based ATS system outperformed the PMA-based system in our experiment, this finding does not negate the merits of PMA technology. Since PMA has shown the abilities of reaching a sufficiently good level in ATS (Gon-

zalez et al., 2014, 2017a,b; Cheah et al., 2018). Therefore, it is still a good fit for SSI application.

In this study, we focused on the comparison of PMA and EMA, and only tongue tip motion was used for ATS performance. Other studies in literature that have incorporated lip motion and other tongue flesh point motion have achieved high performance for PMA-based ATS (Gonzalez et al., 2014, 2017a,b; Cheah et al., 2018). In addition, this study used on MCD as the ATS performance measure. While MCD is a widely used measure for ATS performance, it does not fully represent the vocal quality of the resulting speech. Other acoustic measures including band aperiodicities distortion (BAP) (Morise, 2016), root mean square error of fundamental frequencies (F0-RMSE), and voiced/unvoiced (V/UV) error rate, as well as listening tests are needed to truly assess the differences of PMA and EMA which has not been conducted in the current stage of this study as explained.

Although the subjects were age- and gender-matched in the two groups for comparison (PMA vs EMA) with the same protocol (stimuli and data size), they were different subjects. Indeed, the PMA and EMA systems were located in two different research laboratories, and they could not be placed at a same location for this study. Because the data were collected by two different teams and with different subjects for the EMA and PMA, there could likely be variations in the outcome of the study between the datasets. This issue will be resolved in the future study where the same subjects will use both devices and the same operators will supervise the data collection sessions.

## 5 Conclusion and Future Work

In this study, we compared the ATS performance between a PMA-based tongue motion tracking device and a commercially available EMA (NDI Wave). We found both the raw magnetic signals and transformed positional signals acquired from PMA have similar ATS performance. Although we found that PMA-based system did not perform as well as the EMA-based system in this single-tracer comparison, PMA still has great potential for SSI application, because it is wireless, affordable, portable, and easy to use. Future work will verify these findings using a larger data set (both EMA and PMA) collected from the same speakers, and further improve the PMA measurement

accuracy as well as the localization approach that converts raw magnetic signals to positional data.

## Acknowledgments

This work was supported by the National Institutes of Health (NIH) under award number R03DC013990 and by the American Speech-Language-Hearing Foundation through a New Century Scholar Research Grant. We also thank Dr. Maysam Ghovanloo and the volunteering participants.

## References

- Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. 2016. Tensorflow: A System for Large-Scale Machine Learning. In *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI'16)*, pages 265–283.
- Byron J Bailey, Jonas T Johnson, and Shawn D Newlands. 2006. *Head & Neck Surgery—Otolaryngology*, volume 1. Lippincott Williams & Wilkins.
- David R Baraff. 1994. Artificial Larynx. US Patent 5,326,349.
- Jeffrey J Berry. 2011. Accuracy of the NDI Wave Speech Research System. *Journal of Speech, Language, and Hearing Research*, pages 1295–1301.
- Florent Bocquelet, Thomas Hueber, Laurent Girin, Christophe Savariaux, and Blaise Yvert. 2016. Real-Time Control of an Articulatory-Based Speech Synthesizer for Brain Computer Interfaces. *PLoS computational biology*, 12(11):e1005119.
- Beiming Cao, Myungjong Kim, J van Santen, T Mau, and J Wang. 2018. Articulation-to-Speech Synthesis Using Articulatory Flesh Point Sensors Orientation Information. In *Proc. INTERSPEECH*, pages 3152–3156.
- Lam Aun Cheah, James M Gilbert, José A González, Phil D Green, Stephen R Eil, Roger K Moore, and Ed Holdsworth. 2018. A Wearable Silent Speech Interface based on Magnetic Sensors with Motion-Artifact Removal. pages 56–62.
- Tamás Gábor Csapó, Tamás Grósz, Gábor Gosztolya, László Tóth, and Alexandra Markó. 2017. DNN-Based Ultrasound-to-Speech Conversion for a Silent Speech Interface. *Proc. Interspeech 2017*, pages 3672–3676.
- Bruce Denby, Tanja Schultz, Kiyoshi Honda, Thomas Hueber, Jim M Gilbert, and Jonathan S Brumberg. 2010. Silent Speech Interfaces. *Speech Communication*, 52(4):270–287.

- Lorenz Diener, Sebastian Bredehoeft, and Tanja Schultz. 2018. A comparison of EMG-to-Speech Conversion for Isolated and Continuous Speech. In *Speech Communication; 13th ITG-Symposium*, pages 1–5. VDE.
- Sharon Glennen and Denise C DeCoste. 1997. *The Handbook of Augmentative and Alternative Communication*. Cengage Learning.
- Jose A Gonzalez, Lam A Cheah, Jie Bai, Stephen R Ell, James M Gilbert, Roger K Moore, and Phil D Green. 2014. Analysis of Phonetic Similarity in a Silent Speech Interface Based on Permanent Magnetic Articulography. In *Proc. INTERSPEECH*, pages 1018–1022.
- Jose A Gonzalez, Lam A Cheah, Angel M Gomez, Phil D Green, James M Gilbert, Stephen R Ell, Roger K Moore, and Ed Holdsworth. 2017a. Direct Speech Reconstruction from Articulatory Sensor Data by Machine Learning. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(12):2362–2374.
- Jose A Gonzalez, Lam A Cheah, Phil D Green, James M Gilbert, Stephen R Ell, Roger K Moore, and Ed Holdsworth. 2017b. Evaluation of a Silent Speech Interface Based on Magnetic Sensing and Deep Learning for a Phonetically Rich Vocabulary. *Proc. Interspeech 2017*, pages 3986–3990.
- John C Gower. 1975. Generalized Procrustes Analysis. *Psychometrika*, 40(1):33–51.
- Melvin Hyman. 1955. An Experimental Study of Artificial-Larynx and Esophageal Speech. *Journal of Speech and Hearing Disorders*, 20(3):291–299.
- Myungjong Kim, Beiming Cao, Ted Mau, and Jun Wang. 2017. Speaker-Independent Silent Speech Recognition from Flesh-Point Articulatory Movements Using an LSTM Neural Network. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 25(12):2323–2336.
- Myungjong Kim, Nordine Sebkhi, Beiming Cao, Maysam Ghovanloo, and Jun Wang. 2018. Preliminary Test of a Wireless Magnetic Tongue Tracking System for Silent Speech Interface. In *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pages 1–4. IEEE.
- Diederik P Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980*.
- Anne Laumann, Jaimee Holbrook, Julia Minocha, Diane Rowles, Beatrice Nardone, Dennis West, Jeonghee Kim, Joy Bruce, Elliot Roth, and Maysam Ghovanloo. 2015. Safety and Efficacy of Medically Performed Tongue Piercing in People with Tetraplegia for Use with Tongue-Operated Assistive Technology. *Topics in Spinal Cord Injury Rehabilitation*, 21(1):61–76.
- Ted Mau. 2010. Diagnostic Evaluation and Management of Hoarseness. *Medical Clinics*, 94(5):945–960.
- Ted Mau, Joseph Muhlestein, Sean Callahan, and Roger W Chan. 2012. Modulating Phonation Through Alteration of Vocal Fold Medial Surface Contour. *The Laryngoscope*, 122(9):2005–2014.
- J. Mertl, E. kov, and B. epov. 2018. Quality of Life of Patients After Total Laryngectomy: the Struggle Against Stigmatization and Social Exclusion Using Speech Synthesis. *Disability and Rehabilitation: Assistive Technology*, 13(4):342–352.
- Masanori Morise. 2015. CheapTrick, A Spectral Envelope Estimator for High-Quality Speech Synthesis. *Speech Communication*, 67:1–7.
- Masanori Morise. 2016. D4C, A Band-Aperiodicity Estimator for High-Quality Speech Synthesis. *Speech Communication*, 84:57–65.
- Yoshitaka Nakajima, Hideki Kashioka, Kiyohiro Shikano, and Nick Campbell. 2003. Non-Audible Murmur Recognition Input Interface Using Stethoscopic Microphone Attached to the Skin. In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings (ICASSP'03)*, volume 5, pages 708–711. IEEE.
- Joanne Robbins, Hilda B Fisher, Eric C Blom, and Mark I Singer. 1984. A Comparative Acoustic Study of Normal, Esophageal, and Tracheoesophageal Speech Production. *Journal of Speech and Hearing disorders*, 49(2):202–210.
- Paul W Schönle, Klaus Gräbe, Peter Wenig, Jörg Höhne, Jörg Schrader, and Bastian Conrad. 1987. Electromagnetic Articulography: Use of Alternating Magnetic Fields for Tracking Movements of Multiple Points Inside and Outside the Vocal Tract. *Brain and Language*, 31(1):26–35.
- Nordine Sebkhi, Dhyey Desai, Mohammad Islam, Jun Lu, Kimberly Wilson, and Maysam Ghovanloo. 2017. Multimodal Speech Capture System for Speech Rehabilitation and Learning. *IEEE Transactions on Biomedical Engineering*, 64(11):2639–2649.
- Yana Yunusova, Jordan R Green, and Antje Mefferd. 2009. Accuracy Assessment for AG500, Electromagnetic Articulograph. *Journal of Speech, Language, and Hearing Research*, pages 547–555.