In Submission to *Computer Methods in Biomechanics and Biomedical Engineering*
Vol. 00, No. 00, Month 20XX, 1–11

Electromagnetic Articulography (EMA) for Real-time Feedback Application: Computational Techniques

B. Haworth[a,c*], E. Kearney[b,c], P. Faloutsos[a,c], M. Baljko[a,c], and Y. Yunusova[b,c,d]

[a] *Department of Electrical Engineering and Computer Science, York University, Toronto, Canada*;
[b] *Department of Speech-Language Pathology, University of Toronto, Toronto, Canada*;
[c] *University Health Network: Toronto Rehabilitation Institute, Toronto, Canada*
[d] *Brain Sciences, Sunnybrook Research Institute, Toronto, Canada*

(*Received December 2016*)

The application of state-of-the-art signal processing often differs between off-line and on-line real-time application domains. Off-line processing techniques may be used to accurately reduce signal noise and spot errors before analysis. However, without the global signal information available to off-line processes, such techniques can be difficult to reproduce in on-line real-time applications. This paper presented methods that were developed to support a state-of-the-art Computer-Based Speech Therapy System. These methods included on-line head correction and low-pass filtering and aimed to reproduce off-line processing data quality when using a real-time clinical feedback application. The adequacy of these methods was evaluated relative to the off-line processing "gold" standard and in a context of computing a specific kinematic parameter (i.e. articulatory working space). The results showed that the on-line real-time output values were highly correlated with the off-line manually-processed values.

Keywords: Electromagnetic Articulography (EMA), Wave Speech Research System, Speech Kinematics, Computer-Based Speech Therapy

1. Introduction

The use of augmented kinematic visual feedback for motor learning and recovery has been supported by motor learning and rehabilitation science and practice, fields that are currently moving towards visualization and gamification. In the realm of speech analysis and rehabilitation, research has been mostly concerned with speech acoustics. There is a rapidly growing interest, however, to analyze articulatory kinematics and apply state-of-the-art practices to rehabilitation of motor speech disorders such as dysarthria and apraxia of speech (AOS). It is our current premise that an effective and usable system will translate into meaningful quality-of-life outcomes for many people.

Electromagnetic articulography (EMA) sensor technology holds great potential for new advances in user-oriented health and wellness applications such as speech therapy and accent modification. EMA provides access to the kinematics of articulators such as the jaw, lips, and particularly the tongue, which is typically hidden from view during speech. There are however a number of challenges in employing EMA, including sensor noise, erroneous artifacts, missing data, and necessary data transformations. The standards for addressing these issues in post-processing, however, have been established (Green, Wang and Wilson 2013; Gracco 1992). The real-time on-line processing methods relevant to various clinical applications have not been established. In this paper, we describe and address a series of computational issues concerning the use of EMA sensor technology, as deployed in the specific application domain of computer-based speech therapy (CBST).

*Corresponding author. Email: brandon@eecs.yorku.ca

2. Background

2.1 *Electromagnetic Articulography (EMA)*

EMA is a sensor-tracking technology based on the principles of electromagnetic induction and is a powerful alternative to other articulatory tracking methods such as cineradiography, x-ray microbeam, and ultrasound. The creation of EMA stems from a long history of the need to accurately track articulators during speech and non-speech tasks (Hixon 1971). While visible articulator movements, such as those by the lips and jaw, can be tracked using a variety of both custom- and commercially-developed technologies, tracking the hidden tongue presents challenges.

Early methods for tracking the tongue were primarily limited to two-dimensional (2D) data outputs, and the devices required lengthy calibration processes (Perkell, Cohen, et al. 1992; Schönle, Gräbe, et al. 1987). Later methods afforded full three-dimensional (6D) trackingthe position and rotation of sensors on the tongue (Kaburagi, Wakamiya and Honda 2005; Zierdt 1993). Commercial speech research solutions are now readily available, such as the Carstens AG500 line of products (Carstens Medizinelektronik GmbH, Bovenden) and the Wave Speech Research (NDI, Waterloo) systems. These commercial systems have been tested for their accuracy and demonstrate adequate performance (Savariaux, Badin, et al. 2017; Berry 2011; Yunusova, Green and Mefferd 2009; Kroos 2012). Most of these current commercial systems, such as the Wave, produce audio-aligned six-dimensional (6D) kinematic time series information, and they come with recording and data transformation software, as well as APIs (Application Programmers Interfaces) for developing external applications.

2.2 *Computer-Based Speech Therapy using EMA*

EMA has been successfully deployed in the domain of CBST. This deployment spans the clinical spectrum including speech therapy for accent training, neurologic disorders, and hearing/deafness. For example, Levitt and Katz (2010) reported success when using EMA to facilitate training of a Japanese flap in eight monolingual English speakers. Children with hearing impairment were successfully trained to produce Mandarin words using an EMA-driven “talking-head”; improvements in articulation of bilabial, alveolar, and retroflex consonants with subsequent increases in speech intelligibility were reported post training (Liu, Yan, et al. 2013).

The literature also describes efficacious applications of EMA-provided visual feedback in speech therapy for AOS post stroke, a condition characterized by the inability to achieve consistently correct articulatory positions for speech sounds, resulting in frequent speech errors (Katz and McNeil 2010; Katz, McNeil and Garst 2010). EMA-supplied augmented feedback led to improvements in the accuracy of tongue placement and speaking abilities in a number of speakers with AOS (Katz and Mehta 2015; Katz, McNeil and Garst 2010; Katz, Carter and Levitt 2007; Katz, Bharadwaj, et al. 2002; Katz, Bharadwaj and Carstens 1999).

Recently, OptiSpeech, was designed to deliver positional targets to train articulator accuracy and repeatability of place of articulation for American English consonants (Katz, Campbell, et al. 2014). Our group has previously reported on the design of a CBST system to deliver game-based visualizations to improve speech production for patients with dysarthria due to Parkinson’s Disease (PD) (Yunusova, Kearney, et al. 2017; Haworth, Kearney, et al. 2014; Shtern, Haworth, et al. 2012).

2.3 *Data Quality and Processing*

As therapeutic developments move forward, technological limitations and challenges of the EMA systems have to be carefully considered. A number of existing studies have addressed the issue of quality and accuracy of data captured by EMA devices (Savariaux, Badin, et al. 2017; Berry 2011; Yunusova, Green and Mefferd 2009; Kroos 2008).

The analysis of positional and rotational data from the Carstens AG500 has revealed various data artifacts. A number of these artifacts were dependent on the characteristics of the electromagnetic field. In some active regions, the accuracy of the AG500 worsened, revealing maximum error estimates of up to 5mm for non-speech and 2mm for speech movements (Yunusova, Green and Mefferd 2009; Kroos 2008).

The Wave system's positional error increases relative to the orthogonal distance from the field generator. Estimates made using a rigid four-bar linkage to simulate speech-like dynamic movement reveal that error grows upwards of 9mm and 66mm for the 300mm^2 and 500mm^2 active fields respectively (Berry 2011). In speech movements, the error estimates were relatively small, being controllably sub-millimetre within 150mm of the field generator. The difference in error estimate between the simulated dynamical non-speech movements and speech movements is attributed to slower sensor speeds and smaller ranges of variation in orientation during speech.

Furthermore, a recent comparative study of the Wave system and the AG500 revealed troublesome errors (Savariaux, Badin, et al. 2017). These errors ranged in magnitude from 0.3mm to 21.8mm and increased depending on the head/sensor position relative to the field generator. The Wave was prone to errors that were larger in magnitude than the AG500. In the Wave, the lowest errors were associated with the negative axes closest to the field generator. It is recommended that when using the Wave, the speaker is positioned close to the field generator and oriented the same way as the reference system.

In addition to the errors associated with the field positioning, error-producing issues may include: (1) expected high-frequency noise; (2) sudden rapid positional jumps, or spikes, in data potentially due to electromagnetic noise in the environment; (3) and errors related to rapid changes in, or high variability of, velocity and orientation as noted with artificial movements (see Berry (2011)) but not speech movements (Savariaux, Badin, et al. 2017). In addition to these tracking errors, missing data may occur due to out of field or briefly malfunctioning sensors. In addition to error sources, the freedom-of-movement field-based nature of these types of systems also introduces a need to account for head motions, which is often accomplished by re-orienting articulators sensors to the reference (head) sensor (Perkell, Cohen, et al. 1992; Westbury 1991).

Overall, these issues may lead to erroneous raw and derived data with inflated measurement variability. Only after considering all of these potential sources of error and data variability can we implement EMA for tracking articulatory motion on-line in a therapeutic context. This paper seeks to address these issues via the evaluation of the on-line real-time data processing routines and their effect on the derived measure relative to the existing post-processing "gold" standards.

3. Methods

3.1 *Instrumentation*

Speech movement tracking requirements were realized by the Wave Speech Research System. Our sensor array is composed of (i) a 6 Degree of Freedom (DoF) sensor fixed to the head, and (ii) a single 5 DoF tongue sensor. The head and tongue sensors are shown in Fig. 1. The tongue sensor is attached on the tongue blade, approximately 1cm away from the tip, by means of non-toxic dental glue (PeriAcryl®90, Glustitch). Participants were positioned relative to the field generator so as to reduce tracking errors (Savariaux, Badin, et al. 2017). Movement data were acquired at a sampling rate of 100Hz,

Figure 1. An example sensor setup showing the head sensor, attached to the head strap, and the tongue sensor which is affixed directly to the tongue using dental glue.

which is lower than the highest available sampling rate of the system (400Hz). The lower rate reduces errors associated with data buffering along the communication channels. It is also sufficient for our purposes, since the frequency range of the tongue's motion associated with typical speech production is under 20Hz (Gracco 1992).

The hardware architecture of our system has been described elsewhere (Shtern, Haworth, et al. 2012). Data from the Wave was temporarily stored locally on the Wave control machine for offline processing. Remote real-time data streaming for on-line real time data access was accomplished via the TCP–based Real-Time Application Program Interface (RTAPI) (NDI, Waterloo). The RTAPI allows for custom-built software to access Wave sensor data from a remote or networked computer.

3.2 *Offline Processing*

The standard post-processing routines for kinematic data include: (a) head correction; (b) data re-sampling; and (c) data filtering (Westbury 1994; Gracco 1992). Head correction with the Wave is performed through a black-box export method, provided by software included with the Wave system; it effectively re-expresses data relative to the head-based coordinate system as opposed to the field generator coordinate system. The signal is re-sampled regularly in time using a cubic interpolating spline, since many global smoothing filters assume a regular sampling period and small timing inconsistencies may occur during recording. The data is then smoothed using a low-pass filter (Green, Wang and Wilson 2013). The filter has an empirically determined cut-off frequency depending on the articulator (e.g., jaw versus tongue tip versus tongue dorsum) and phonetic context analyzed (Gracco 1992). We used a 15Hz cut-off frequency for tongue blade movement data.

3.3 *On-Line Processing Pipeline*

An on-line processing pipeline has been developed to rectify the data during real time acquisition. The pipeline consists of two main processes employed prior to derivation of necessary kinematic measures or metrics (Section 3.4) - head correction then filtering.

3.3.1 Head Correction

The head correction transformation is predicated on a head position vector and rotation quaternion, p_h, q_h respectively. The tongue sensor, or any other sensor to be head corrected, is represented by a position vector and an orientation quaternion, p_t, q_t respectively. Finally, the head corrected tongue position and rotation, p_t^h, q_t^h, are computed as follows:

$$p_t^h = Im(q_h^{-1} * (0, p_t - p_h)), \qquad q_t^h = q_h^{-1} * q_t \tag{1}$$

where Im is the imaginary part of a quaternion, and '*' indicates quaternion multiplication. In practice, to subtract head rotations, an angle axis rotation is formed by the quaternion and the sensor is rotated around the point p_i after being translated by $p_t - p_h$.

3.3.2 Filtering

To address the variety of EMA-based error sources in a real-time setting, noted in Section 2.3, a moving median filter is employed. A moving median filter is a standard method of low pass filtering data (Justusson 1981). The moving median works locally on the filter window and handles rapid artifacts while avoiding the introduction of artificial data - as may occur with averaging filters. The window sizes were determined empirically to minimize processing time (median filters must sort data first), and qualitatively reduce high frequency noise while preserving known speech movement features. A window size of 3-5 samples were determined to work best for data sampled at 100Hz, while a 6-21 sample window size worked best for data sampled at 400Hz.

3.4 *Metric Derivation: Articulatory Working Space (AWS)*

A variety of kinematic metrics can be derived from EMA-based data. The choice of metric is highly dependent on the treatment population and the focus of the therapy. For example, Katz and colleagues used an experimenter-defined circular target region at the alveolar ridge to train elevation during consonant sounds in speakers with AOS (Katz and Mehta 2015; Katz and McNeil 2010; Katz, Carter and Levitt 2007).

The pilot target population for our initial set of studies were adults diagnosed with a speech disorder (e.g. dysarthria) due to PD. This population shows an overall reduction in articulatory movements in the lips, jaw, and tongue during speaking (Walsh and Smith 2012; Weismer, Yunusova and Bunton 2012). This reduction of movement size is reflected in the individual's articulatory working space (AWS, a 2D representation of which is shown in Figure 2), the convex hull surrounding the movement trajectory traversed during a speaking task (Weismer, Yunusova and Bunton 2012). The spatial volume (mm^3) of the 3D AWS is used to characterize a patient's movement range. Thus, increase in AWS over the course of therapy, is chosen as a treatment target and yoked to real-time visual feedback in a CBST system.

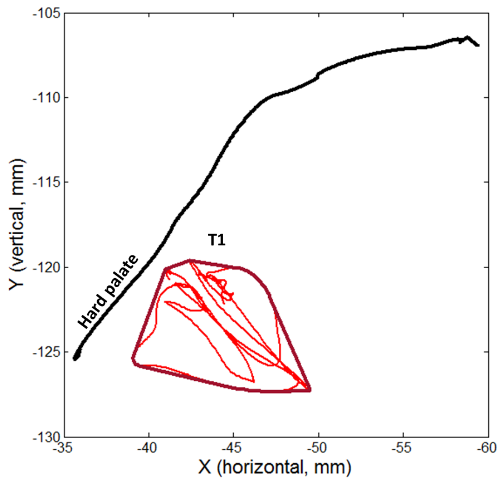

Figure 2. An illustrative AWS for tongue blade (T1) sensor, when tongue movement was generated by a male speaker with PD saying the sentence, "Sally sells sevens spices."

3.4.1 *Real-time AWS Volume*

A 3D convex hull around articulator trajectories results in an irregular polyhedron. To discretize the space of this hull and compute its volume, i.e. operationalize the AWS, a Delaunay triangulation in three dimensions, or tetrahedralization is found. This algorithm generates space-filling tetrahedrons whose combined free surface forms a convex hull. The volume can then be computed as the total sum of the tetrahedron volumes, as follows:

$$V_{AWS} = \sum_{t \in T} \frac{|(a_t - d_t) \cdot ((b_t - d_t) \times (c_t - d_t))|}{6} \tag{2}$$

where $\cdot$ and $\times$ are the dot and cross product respectively, a_t, b_t, c_t, d_t are the vertices of the tetrahedron t, and T is the set of all tetrahedrons.

The Delaunay triangulation algorithm used in this work is a real-time approach inspired by the classic QHull algorithm (Sehnal and Campbell 2014). Since QHull-based convex hulls may "collapse" under certain degenerate point additions, the real time AWS derivation is sensitive to degenerate regular point configurations (colinear, cospherical, coplanar, and grids), which must be taken into account while computing AWS. For real-time purposes, degenerate points are dealt with using pairwise identical point removal and input joggling. The pairwise removal operation culls points considered equal in position, i.e. when the distance between the two is less than an epsilon value (10^{-7}mm). The input joggling process is carried out by point vector addition of random noise within a sphere of radius small enough not to impact measures (10^{-6}mm).

3.5 *Evaluation*

Three analyses were conducted, comparing the on-line with off-line ("gold" standard) processing approaches. These included: (1) head correction; (2) low-pass filtering; and (3) metric derivation are presented below.

3.5.1 Comparison of Head Correction Routines

To test the accuracy of head correction between the on-line and off-line methods, a single 6 DoF and one 5 DoF sensor were fixed to a rigid wooden splint 50mm apart. The 6 DoF sensor was used as a reference system for the 5 DoF sensor.

Task. Four separate tests were conducted using the rigid body configuration: (1) near field generator stationary sensors test (near-static); (2) near field generator moving sensors test (near-moving); (3) far field generator static sensors test (far-static); (4) and far field generator moving sensors test (far-moving). The following distances were measured orthogonal to the patient-facing side of the field generator. The static sensor tests were conducted at a fixed distance, with the near experiment at 100mm from the field generator and the far experiments 200mm from the field generator. The moving sensor tests were conducted by making random translational and rotational movements within 150mm from the field generator for the near experiments and further than 150mm for the far experiments.

Measure. The head-corrected positional data for the 5 DoF sensor was recorded using each of the on-line and the off-line pipelines. For each pipeline condition, the standard deviation of the values in each of the three positional dimensions (X, Y, & Z) was derived. Flawless head correction would produce a value of 0 in each dimension.

3.5.2 Comparison of Filtering Routines

To test the effect of two different filtering approaches on tongue kinematics, data were collected from a single speaker (Male, 23 years of age). Tongue blade movements were captured with a single 5 DoF sensor in real time and head corrected using the built-in Wave routine before being median filtered (on-line) and low-pass filtered (off-line) using a bi-directional low-pass 5th-order Butterworth filter.

Task. The participant was asked to repeat a list of 37 different sentences at a normal comfortable speaking rate and loudness. Only recordings where zero positional data loss occurred ($N = 25$) were used in the analysis.

Measure. Kinematic data, head-corrected and filtered using the respective on-line versus off-line routines, were compared by measuring the Root-Mean-Square Error (RMSE) between the two processing sources.

3.5.3 Comparison of Metric Computation Routines

To compare the on-line and off-line derived AWS metrics, tongue blade movements were collected from a large set of clinical participants.

Participants. Nine participants were recruited for a study of articulatory movements in PD. The group included seven males and two females between the ages 57 and 90 diagnosed with PD and at various times post-diagnosis. All participants reported to be optimally medicated and not fatigued before the recording session (Fisk and Doble 2002). The primary inclusion criteria were the clear presence of hypokinetic dysarthria with impairment of speech intelligibility and perceptual deficits in the articulatory domain (i.e., imprecise consonants, distorted vowels, and short rushes of speech). All participants provided informed consent and were covered under the University Health Network Research Ethics Board (*reference: 13-6235-DE*)

Task. Each participant produced on average fifteen repetitions (9 - 20 per participant) of 4 sentences in a random order ($N = 742$ total).

Measures. Kinematics signals, head-corrected and filtered using the respective on-line versus off-line routines, were used to compute the AWS as well as the duration of each sentence. Sentence durations were recorded, as a difference between the timing of speech onsets and offsets was expected between the on-line processed and off-line post-processed recordings. This difference reflected the fact that during the on-line and off-line procedures, the onsets/offsets were controlled manually and separately by two different human operators. On-line processed recordings were segmented by a clinician (record start/stop) who was performing a CBST session, however the off-line processed recordings were segmented by an expert operator (research assistant) during post-processing. To ensure fair comparisons the recordings with differences in durations between the two methods, larger than 0.5 seconds, have been removed from further analysis ($N = 255$ total). As a result, a total of 487 recordings were analyzed.

3.6 *Results*

3.6.1 Head Correction

Head correction results are summarized in Figure 3. The head-corrected 5 DoF sensor positions, using on-line and off-line head correction methods, revealed that the error (standard deviations) were relatively small, particularly for near static and moving conditions. Far-field conditions provided larger deviations, which were particularly notable in the moving condition. This was likely due to the increases in error associated with orthogonal distance from field generator, previously reported in the literature. Interestingly, the far-moving condition showed more variability with the off-line as compared to the on-line method. A Levene's test (Levene, et al. 1960), to understand the homogeneity of variance, showed that the variance was significantly different between on-line and off-line head-correction methods across all conditions ($p < 0.0001$). The difference between on-line and off-line methods are difficult to fully understand because the off-line method for head correction is closed-source and unavailable for further analysis.

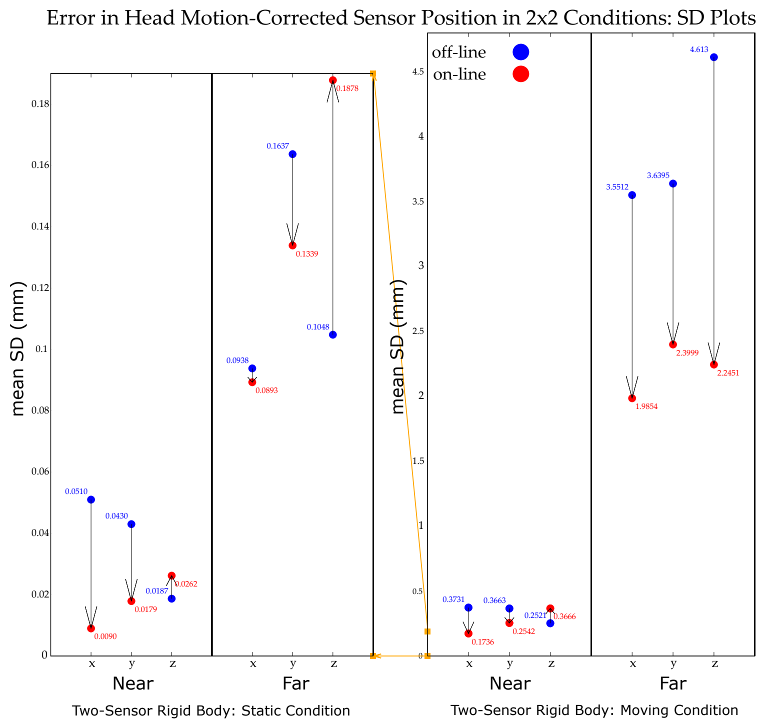

Figure 3. Standard deviations, in mm, of a rigid body 5 DoF sensor after head correction obtained using the on-line and off-line pipelines. The orange line shows the scaling factor of the static sensor error, as static sensor errors were orders of magnitudes smaller than moving sensors.

3.6.2 Filtering

Figure 4 shows an example of the filtering process effects when the kinematic signal is noisy. Summary statistics across sentences indicated that the median and low-pass filtering methods produced comparable signals, when RMSE for each dimension (X, Y, Z) was measured between the output signals of each filtering method. The mean RMSE and standard deviation values were: X - 0.1883 ± 0.0745mm; Y - 0.1545 ± 0.0802mm; and Z - 0.2702 ± 0.1376mm. These RMSE values suggested that the off-line filtering approach was well within the experimental bounds required by the application of real-time feedback.

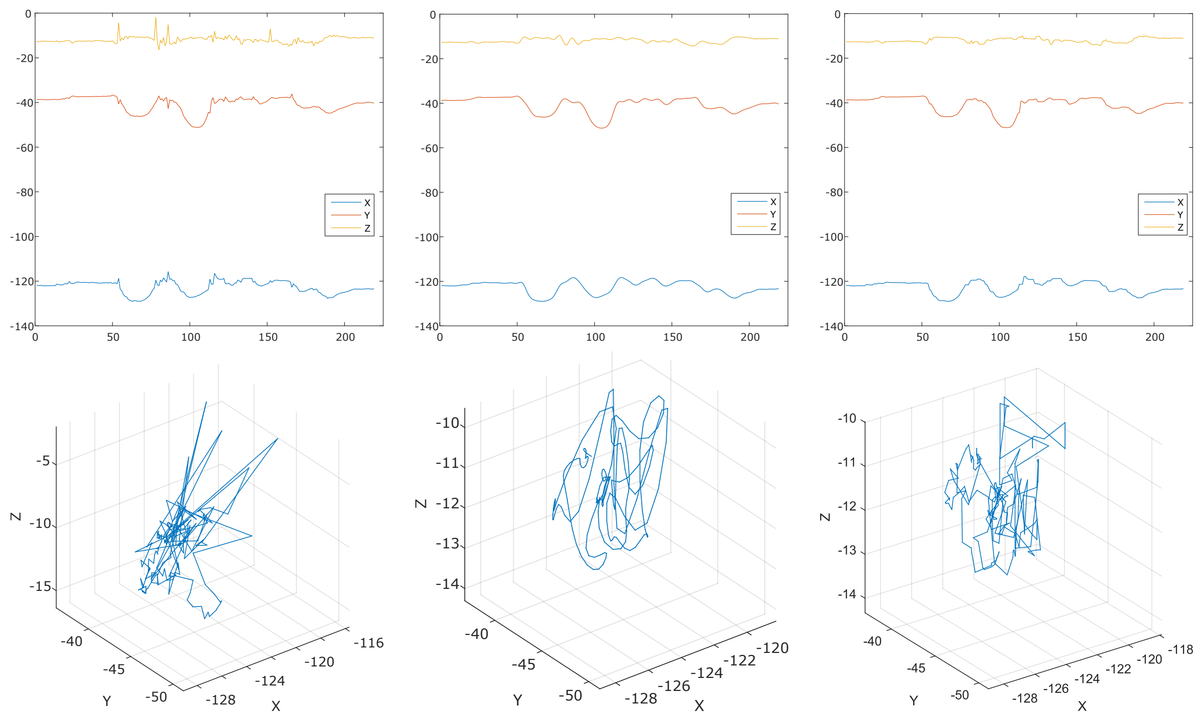

Figure 4. An example of the effects of filtering method on a tongue movement signal recorded during the sentence, "Tom took the tasty teas on the terrace", with (a) showing the raw signal, (b) the off-line filtered signal, and (c) the on-line filtered signal. The top row is the signal decomposed into the three dimensions (X, Y, Z) and the bottom row is its 3D trajectory. The high frequency noise seen in (a) is removed by both methods. The range of the axes in (a) is greater due to noisy spikes.

3.6.3 Metric Computation

The AWS values obtained during off-line and on-line computations were compared using a Bland-Altman analysis approach (Bland and Altman 1999, 1986). Figure 5 shows the typical correlation and difference plot associated with the Bland-Altman analysis.

The coefficient of determination for the two methods was 0.9082, and visual analysis revealed that the majority of samples had excellent agreement. Given the high correlation, the sum of squared errors of prediction was within acceptable limits for real time feedback ($< 150\text{mm}^3$).

4. Conclusion

In summary, while addressing challenges associated with data acquisition using the Wave system, we compared two methods of data post-processing and metric derivation - an offline "gold" standard and an on-line method developed in-house for a specific real-time data streaming purpose. Overall, the on-line procedures were comparable to the off-line procedures. These results demonstrated

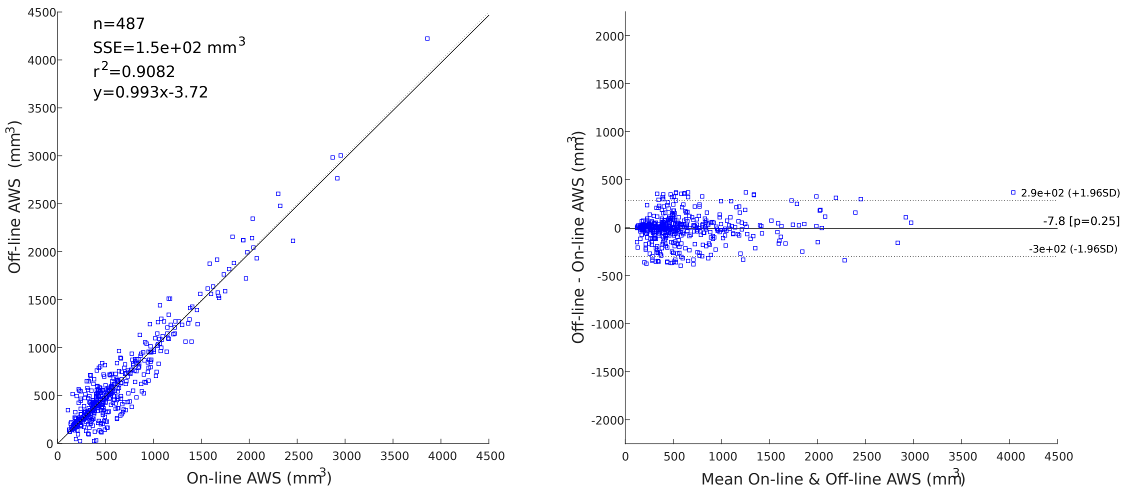

Figure 5. Bland-Altman plots comparing on-line to off-line AWS derivation.

that we can derive various metrics, AWS and beyond, to characterize movements of the speech articulators in real time and use this information for the development clinical applications that require real-time or near real-time data display. The techniques presented here ensure reliable derivation of these measures in an automatic and operator independent manner.

Both clients and clinicians would benefit from a system that affords augmented feedback through movement visualization while providing the underlying computational requirements of an experimental rehabilitation framework. The aforementioned techniques have been instantiated computationally in a research prototype and experimental apparatus. This apparatus has been deployed to investigate the impact of various visual feedback factors on clinical outcomes (Yunusova, Kearney, et al. 2017).

Limitations in the current study are left for future work. These include an in-depth analysis of the current state-of-the-art offline filtering methods, missing samples reconstruction, and the decision boundary for reconstructing or discarding data.

5. Acknowledgements

This research has been supported from multiple sources: the NSERC Discovery Grant Program, the ASHA Foundation (New Investigator Award), the University Health Network—Toronto Rehabilitation Institute, the Parkinson Society of Canada Pilot Project Grant Program, and the Centre for Innovation in Information Visualization and Data-Driven Design (CIV-DDD).

References

Berry JJ. 2011. Accuracy of the NDI wave speech research system. Journal of Speech, Language, and Hearing Research. 54(5):1295–1301.

Bland JM, Altman D. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet. 327(8476):307–310.

Bland JM, Altman DG. 1999. Measuring agreement in method comparison studies. Statistical Methods in Medical Research. 8(2):135–160.

Fisk JD, Doble SE. 2002. Construction and validation of a fatigue impact scale for daily administration (d-fis). Quality of Life Research. 11(3):263–272.

Gracco VL. 1992. Analysis of speech movements: practical considerations and clinical application. Haskins Laboratories status report on speech research SR-109/110:45–58.

Green JR, Wang J, Wilson DL. 2013. Smash: a tool for articulatory data processing and analysis. In: Interspeech. p. 1331–1335.

Haworth MB, Kearney E, Baljko M, Faloutsos P, Yunusova Y. 2014. Electromagnetic articulography in the development of serious games for speech rehabilitation. Proceedings of the 2nd International Workshop on Biomechanical and Parametric Modeling of Human Anatomy (PMHA 2014).

Hixon TJ. 1971. An electromagnetic method for transducing jaw movements during speech. The Journal of the Acoustical Society of America. 49(2B):603–606.

Justusson B. 1981. Median filtering: Statistical properties. In: Two-dimensional digital signal prcessing ii. Springer; p. 161–196.

Kaburagi T, Wakamiya K, Honda M. 2005. Three-dimensional electromagnetic articulography: A measurement principle. The Journal of the Acoustical Society of America. 118(1):428–443.

Katz W, Bharadwaj S, Carstens B. 1999. Electromagnetic articulography treatment for an adult with Broca's aphasia and apraxia of speech. Journal of Speech, Hearing, and Language Research. 42:1355–1366.

Katz W, Bharadwaj S, Gabbert G, Stettler M. 2002. Visual augmented knowledge of performance: Training place of articulation errors in apraxia of speech using EMA. Brain and Language. 83:187–189.

Katz W, Campbell TF, Wang J, Farrar E, Eubanks JC, Balasubramanian A, Prabhakaran B, Rennaker R. 2014. Opti-speech: A real-time, 3D visual feedback system for speech training. In: Proc. Interspeech.

Katz W, Carter G, Levitt J. 2007. Treating buccofacial apraxia using kinematic feedback. Aphasiology. 21:1230–1247.

Katz W, McNeil M. 2010. Studies of articulatory feedback treatment for apraxia of speech (AOS) based on electromagnetic articulography. Perspectives on Neurophysiology and Neurogenic Speech and Language Disorders. 20(3):73–80.

Katz W, McNeil M, Garst D. 2010. Treating apraxia of speech (AOS) with EMA-supplied visual augmented feedback. Aphasiology. 24(6–8):826–837.

Katz WF, Mehta S. 2015. Visual feedback of tongue movement for novel speech sound learning. Frontiers in Human Neuroscience. 9:612.

Kroos C. 2008. Measurement accuracy in 3D electromagnetic articulography (Carstens AG500). In: Proceedings of the 8th international seminar on speech production. p. 61–64.

Kroos C. 2012. Evaluation of the measurement precision in three-dimensional electromagnetic articulography (Carstens AG500). Journal of Phonetics. 40(3):453–465.

Levene H, et al. 1960. Robust tests for equality of variances. Contributions to Probability and Statistics. 1:278–292.

Levitt JS, Katz WF. 2010. The effects of EMA-based augmented visual feedback on the english speakers acquisition of the Japanese flap: a perceptual study. Stroke. 4:5.

Liu X, Yan N, Wang L, Wu X, Ng ML. 2013. An interactive speech training system with virtual reality articulation for mandarin-speaking hearing impaired children. In: Information and Automation (ICIA), 2013 IEEE International Conference on. IEEE; p. 191–196.

Perkell JS, Cohen MH, Svirsky MA, Matthies ML, Garabieta I, Jackson MT. 1992. Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. The Journal of the Acoustical Society of America. 92(6):3078–3096.

Savariaux C, Badin P, Samson A, Gerber S. 2017. A comparative study of the precision of Carstens and Northern Digital Instruments electromagnetic articulographs. Journal of Speech, Language, and Hearing Research. 60(2):322–340.

Schönle PW, Gräbe K, Wenig P, Höhne J, Schrader J, Conrad B. 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.

Sehnal D, Campbell M. 2014. Miconvexhull library, version "1.0.10.1021". Available from: `https://designengrlab.github.io/MIConvexHull/`.

Shtern M, Haworth M, Yunusova Y, Baljko M, Faloutsos P. 2012. A game system for speech rehabilitation. In: Kallmann M, Bekris K, editors. Motion in games. Lecture notes in computer science; vol. 7660. Springer Berlin Heidelberg; p. 43–54.

Walsh B, Smith A. 2012. Basic parameters of articulatory movements and acoustics in individuals with Parkinson's disease. Movement Disorders. 27(7):843–850.

Weismer G, Yunusova Y, Bunton K. 2012. Measures to evaluate the effects of DBS on speech production. Journal of Neurolinguistics. 25(2):74–94.

Westbury JR. 1991. The significance and measurement of head position during speech production experiments using the x-ray microbeam system. The Journal of the Acoustical Society of America. 89(4):1782–1791.

Westbury JR. 1994. On coordinate systems and the representation of articulatory movements. The Journal of the Acoustical Society of America. 95(4):2271–2273.

Yunusova Y, Green JR, Mefferd A. 2009. Accuracy assessment for AG500, electromagnetic articulograph. Journal of Speech, Language, and Hearing Research. 52:547–555.

Yunusova Y, Kearney E, Kulkarni M, Haworth B, Baljko M, Faloutsos P. 2017. Game-based augmented visual feedback for enlarging speech movements in Parkinson's disease. Journal of Speech, Language, and Hearing Research. 60(6S):1818–1825.

Zierdt A. 1993. Problems of electromagnetic position transduction for a three-dimensional articulographic measurement system. Forschungsberichte-Institut für Phonetik und Sprachliche Kommunikation der Universität München. 31:137–141.