Evaluation of a Wireless Tongue Tracking System on the Identification of Phoneme Landmarks

Nordine Sebkhi[1], Nina Santus[2], Arpan Bhavsar[1], Shayan Siahpoushan[1], and Omer T. Inan[1], *Senior Member, IEEE*

Abstract*—*Objective: **Evaluate the accuracy of a tongue tracking system based on the localization of a permanent magnet to generate a baseline of phoneme landmarks. The positional variability of the landmarks provides an indirect measure of the tracking errors to estimate the position of a small tracer attached on the tongue. The creation of a subject-independent (universal) baseline was also attempted for the first time.** ***Method:*** **2,500 tongue trajectories were collected from 10 subjects tasked to utter 10 repetitions of 25 phonemes. A landmark was identified from each tongue trajectory, and tracking errors were calculated by comparing the distance of each repetition landmark to a final landmark set as their mean position.** ***Results:*** **In the subject-dependent baseline, the tracking errors were found to be generally consistent across all phonemes and subjects, with less than 25% of the errors reported to be greater than 5.8 mm (median: 3.9 mm). However, the inter-subject variability showed that current limitations of our system resulted in appreciable errors (median: 55 mm, Q3: 65 mm).** ***Conclusion:*** **The tracking errors reported in the subject-dependent case demonstrated the potential of our system to generate a baseline of phoneme landmarks. We have identified areas of improvement that will reduce the gap between the subject-dependent and universal baseline, while lowering tracking errors to be comparable to the gold standard.** ***Significance:*** **Creating a baseline of phoneme landmarks can help people affected by speech sound disorders to improve their intelligibility using visual feedback that guides their tongue placement to the proper position.**

***Index Terms*—articulograph, machine learning, permanent magnet localization, phoneme, speech sound disorder, tongue tracking.**

I. Introduction

PHONEMES are one of the elementary building blocks for the production of speech and are the result of a complex mechanism that involves the motion and coordination of the articulators (i.e. tongue, lips, jaw), the vocal folds, and breathing patterns. The intricate mechanism of phoneme production is hindered for individuals affected by speech sound disorders (SSD) which are characterized by diminished motion planning capabilities (apraxia of speech) and/or reduced range of motion due to muscle weakness of the articulators (dysarthria) [1]. In the United States alone, 7.5 million people and nearly 1 out of 12 children [2] are diagnosed with SSD, resulting in a reduction of their ability to communicate effectively. This has negative consequences for their quality of life because many daily activities involve verbal communication. In an attempt to improve speech intelligibility, patients affected by SSD receive therapy services provided by a speech-language pathologist (SLP) whose main objective is to correct the articulators' placement and motion to produce proper sounds. Among all articulators, the tongue is the most important for speech productions [3] but it is also the most difficult to see during speech as it is hidden inside the oral cavity at times during production. This visualization difficulty is a significant issue because it also prevents SLPs from modeling the proper tongue placement to their patients, in the same way they are able to model the lips and jaw that are clearly visible.

As an indirect method to guide patients towards proper tongue placement, SLPs rely on tactile markers to indicate a target landmark for a specific phoneme. Examples of such markers include: tongue depressors, straws, a gloved finger, tongue models, and even flavored lollipops in pediatric practices [4]. Unfortunately, traditional treatment methods that rely on these tools can fail to deliver perceivable progress in proper sound production, as reported with /r/ sounds in children [5]. Additionally, the SLPs assessment of improper tongue placement during therapy can be inaccurate and subjective because s/he has no access to a visualization of how the patient's tongue is moving during examination or while practicing. In short, an SLP cannot see the patient's tongue placement during speech and the patient cannot see the SLP's tongue when modeling proper placement. Furthermore, SLPs only have access to general textbook descriptions of phoneme landmarks which might not be reflective of their actual positions in individual patients.

Therefore, access by SLPs and patients to a real-time visualization of tongue motion during speech would be beneficial during the treatment of SSD. Previous studies [6], [7] have shown that providing such visualization combined with overlaid visual targets for tongue placement can improve the quality of the produced sounds. Aligned with these findings, the overall objective of this research is to develop such visual feedback through the gamification of tongue placement. As illustrated in Fig. 1 (bottom-left), the patient will be able to compare the phoneme landmarks (shown as flowers) to the current placement of her tongue (shown as a bee), and a score will be generated based on the distance errors between them. This score will serve as an objective measure of speech performance which could be used to reinforce good practice and result in improved recovery. To develop this visual feedback, a tongue tracking system is needed not only to provide the tongue position in real-time but also to find where the phoneme landmarks are located. Because of the challenges inherent to

[1]School of Electrical & Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30308, USA.

[2]Department of Communication Sciences and Special Education, University of Georgia, Athens, GA 30602, USA.

Corresponding author: sebnor31@gatech.edu

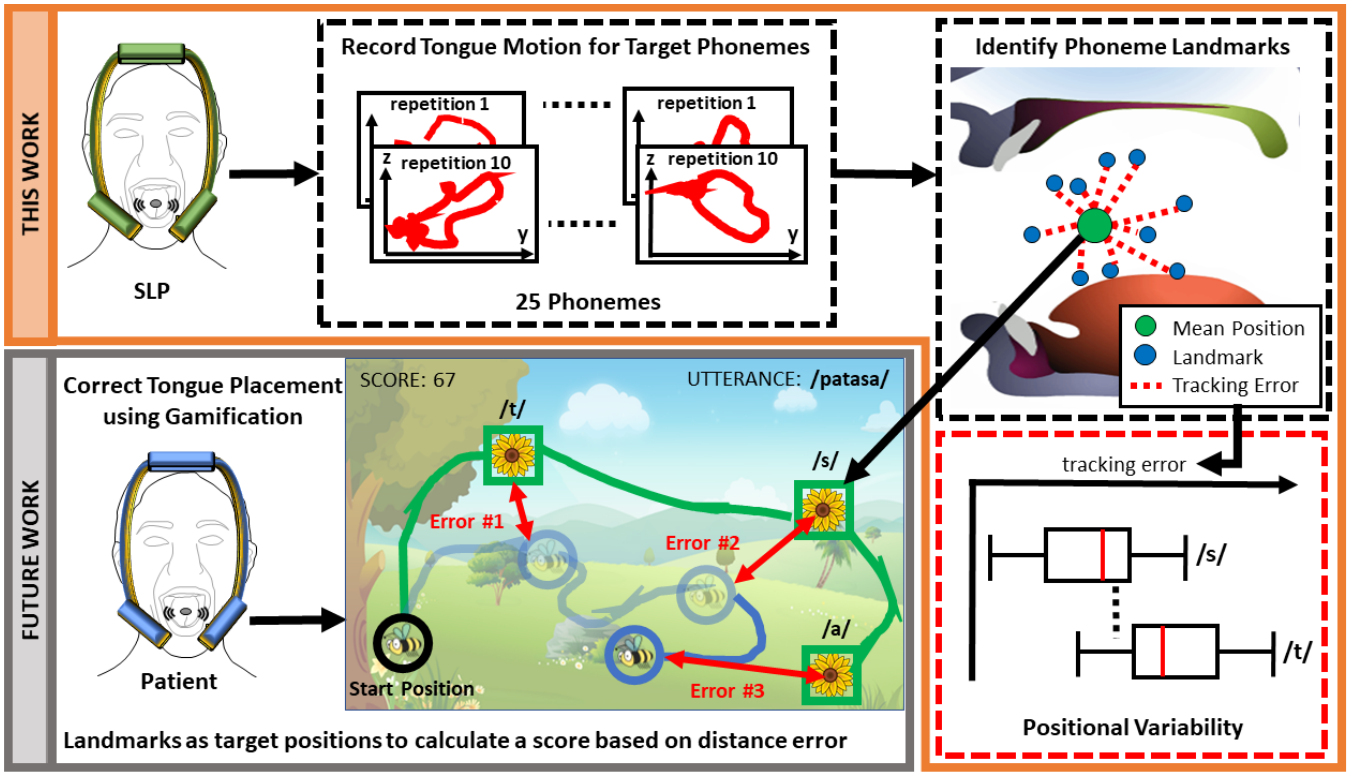

Fig. 1. Overview of the research objective with this work being focused on evaluating the tracking error of the tongue tracking system during speech by calculating error distances from phoneme landmarks identified from tongue motion. These landmarks will then serve as a baseline and display in a game as visual targets to help patients with SSD to better correct their tongue placement.

tracking the tongue, few solutions are available in the market and only two systems are used with regularity by researchers in speech science: the electromagnetic articulograph (EMA) that tracks the motion of multiple points on the tongue using wired sensor probes [8], [9], and the electropalatograph (EPG) that detects points of contact between the tongue and the palate thanks to electrodes embedded in a wired over-the-palate mouthpiece [10], [11]. Although EMA is the gold standard in research because of high tracking accuracy, the tongue probes are wired which are not only an impediment to natural speech [12] but also require a well-trained therapist to be safely attached, and its cost (>$40,000) is significantly beyond the budget of SLP practices. Conversely, EPG is more affordable and is being used by a limited number of SLPs to provide such visual feedback, but it is restricted to the few phonemes that can be produced by tongue-palate contact patterns and the wired mouthpiece is obtrusive which may impede the natural motion of the tongue.

To address some of the aforementioned shortcomings, a tongue tracking system has been developed that (1) minimally impedes natural tongue motion by using a wireless tracking method based on the localization of a permanent magnet attached on the tongue (referred in this paper as tracer), (2) is affordable by using mass-produced components, and (3) can track the tongue anywhere in the oral cavity. The basic functioning of this system is to estimate the 3D position of the tracer from its magnetic field as measured by an array of magnetometers. By affixing the tracer on a desired location on the tongue, this system allows the motion of this marked location to be tracked in real-time and wirelessly. Although the orientation of the tracer can also be tracked, the focus for this study is on the position of the tongue since it has more useful information for speech. In earlier work, the tracking accuracy of the first prototype was assessed in an in-lab test setup [13] and a feasibility study on generating a baseline of phoneme landmarks was conducted [14]. The knowledge obtained from these studies led to the creation of a new version of the system [15] in which the main changes included a complete redesign of the body in the form of a headset to enable the system to be wearable and reduce head/body motion artifacts, and the implementation of a more robust tracking algorithm utilizing deep learning of a feedforward neural network. However, the assessment of the tracking accuracy was performed using an in-lab test setup that emulated tongue motion thanks to a positional stage that accurately placed the tracer at desired, and thus known, positions [15].

To generate a baseline of phoneme landmarks, the tracking accuracy for actual, rather than emulated, tongue motion must be assessed. This is challenging because there is no obvious ground-truth to compare against. In [16], [17], the researchers used an EMA system and attached two probes on the buccal surface of the jaw and at a known distance. The subjects were asked to read a paragraph while the positions of the probes were recorded. The actual distance between the probes remained constant throughout the recording session, thus the tracking error was assessed as its deviation when

compared to the measured distances. This method is well suited to indirectly assess tracking accuracy during speech, but it requires at least two probes to be tracked simultaneously while only one tracer can be tracked in our system design as a trade-off to allow the wireless tracking method. Therefore, our research team constructed a different method to estimate the tracking accuracy of our system during speech.

In this paper, the evaluation of tracking accuracy of our system during actual speech is performed indirectly by generating a baseline of phoneme landmarks. Using the assumption that a phoneme landmark for a person without any SSD is always located at a same position for that individual, any positional variability in tongue placement around that landmark can thus be interpreted as a tracking error. Indeed, if a speaker utters a same phoneme with many repetitions, its positional variability provides a measure of how far apart the landmarks are, and consequently, an indirect measure of how accurate our system is at tracking tongue motion. In reality, the assumption that a phoneme landmark is always located at the same position is not valid because there is a natural positional variability when we produce a phoneme [18], [19]. Therefore, the variability measured with this method is effectively an *upper bound* for the system's tracking error since it also includes this natural tongue placement variability that, although unknown, is supposed to be roughly within a range of 1-3 mm [18].

To generate such a baseline, a human study was conducted to record the tongue motion of 10 SLP students, recruited from the Communication Sciences and Disorders (CMSD) major, that were asked to utter 10 repetitions of a set of 25 phonemes found in American-English and for which tongue placement is critical for their production. As illustrated in Fig. 1, for each phoneme uttered by a subject, a tracking error was calculated between each of the 10 landmarks (one per repetition) and their mean position. These landmarks were also used to generate a universal (i.e. subject-independent) baseline.

The rest of the paper is organized as follows: Section II provides more details about the system, data collection, and analysis; Section III provides the positional variability of the phoneme landmarks; Section IV discusses these results; Section V concludes on the significance of our results and broader impact of this work.

II. Methods

Fig. 1 provides an overview of the method used to evaluate the tracking accuracy of the system. The tongue motion of 10 subjects were recorded wirelessly while each subject produced a set of 25 phonemes for 10 repetitions. Each tongue trajectory was processed to identify a phoneme landmark, and the resulting 2,500 landmarks were analyzed to calculate the tracking errors. More details are provided in the following sections.

A. Tongue Tracking System

Fig. 2 provides an overview of the tongue tracking system and more details can be found in [15]. The tracer generates a magnetic field that is measured by an array of magnetometers at a sampling frequency of 100 Hz. The cylindrical tracer is

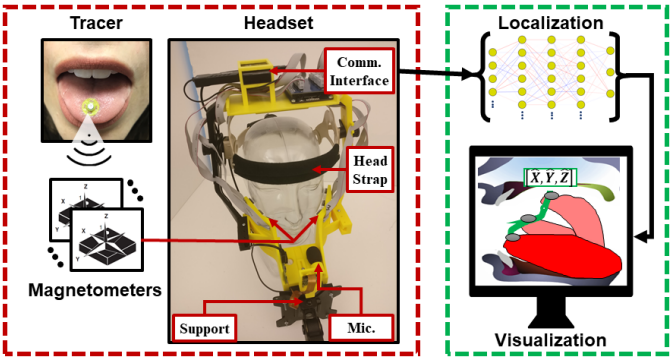

Fig. 2. The system tracks the 3D position of a magnetic tracer attached on the tongue. This wireless tracking relies on a localization algorithm, based on a neural network, that converts the measurements of the tracer's magnetic field by an array of magnetometers to its position. The magnetometers are embedded in a custom-designed headset that transmits all recorded data to a computer using a USB connection.

non-toxic, has a diameter of 4.8 mm and a thickness of 1.6 mm, and is attached on the tongue's blade using an FDA-approved oral adhesive (PeriAcryl®, GluStitch Inc., Canada). The magnetometers are embedded in a stationary headset, designed and 3D-printed in the researchers' lab, and are placed on each side of the mouth. Additionally, the headset houses a microphone for voice recording (96 kHz), and a communication interface to transmit data (i.e. magnetic and voice) to a computer using a standard USB connection. The headset is safely secured on the subject's head thanks to an adjustable strap, and is tethered to an external support to prevent undesired head/body motion. Although this system has the capability to track lip motion, this feature is not used in this study.

In earlier work [15], a new localization algorithm was developed based on a feedforward neural network that mapped the magnetic field read by the magnetometers to the 3D position of the tracer. Our localization algorithm was trained by collecting 1.7 million samples where each sample is a unique combination of 3D positions and 2D rotations of the tracer, and enclosed in a 10x10x10cm^3 volume chosen because that dimension encompasses all sizes of human oral cavities. The median positional error of the localization was reported to be 1.4 mm. The innovative approach of this localization method is that the tracking is performed wirelessly since no wires nor electronics are present in the mouth which enables the user to speak more naturally than other existing tongue tracking systems (e.g. EMA, EPG).

B. Data Collection

Ten CMSD students at the University of Georgia were recruited based on the following criteria: having passed a phonetics class, having no history of speech disorders, and having no intra-oral magnetic device that would interfere with the magnetometers. All subjects were female, between the age of 21-36 y.o., and were raised in the state of Georgia (USA) except for subject #10 from Chicago (Illinois, USA). This study was approved by the Georgia Tech Institutional Review Board and carried out at the University of Georgia.

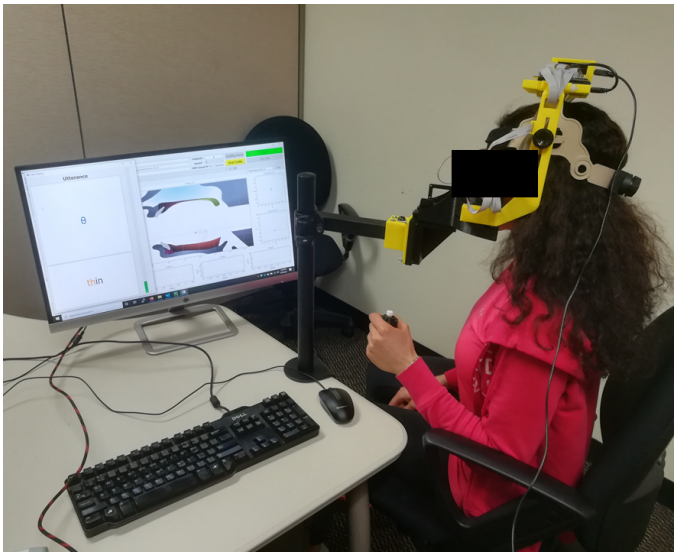

Fig. 3. Illustration of a subject using our tongue tracking system during a data collection session.

A new and sterilized tracer was used for each subject and placed mid-line on the blade of the tongue using the adhesive (~1 cm from the tip). The subjects were first asked to read the Grandfather Passage [20] to become accustomed to speaking with the tracer, and then asked to repeat each phoneme in the following lists with 10 repetitions:

- r sound: [ɔr, aɪr, ar, aʊr, ɪr, ɛr]
- vowel: [æ, ɔ, e, ə, ɚ, ɛ, i, ɪ, u,ʊ]
- consonant: [d, l, n, s, t, z, ʃ, ʒ, θ]

The phonemes were produced in isolation except for the consonants that were followed by a vowel. The subject's voice was also recorded by the headset's built-in microphone, and a clicker was used by the subject to start/stop the recording of each repetition. The subjects were instructed to elongate and articulate their speech, and were allowed to record back any repetition if they feel it was said in error. In an attempt to minimize variation in phoneme production between subjects, a reference audio file was played before each new phoneme and a word was displayed to provide a context in which the target phoneme is used. At the end of the recording session, the tracer was detached from the tongue and disposed. An illustration of a subject using the tongue tracking system during a data collection session is shown in Fig. 3.

C. Data Analysis

The data analysis is roughly composed of three steps (Fig. 4) with each step explained in more details in the following subsections.

1) Landmark Identification: The first step is to identify a landmark for each of the 2,500 tongue trajectories collected in this study (10 subjects x 25 phonemes x 10 repetitions). First, a raw tongue trajectory from our database is extracted for a given subject, phoneme, and repetition. Then, by using its associated audio recording, the period of active speech is identified and the non-speech parts of the raw trajectory are trimmed out. This intermediate step is designed to facilitate

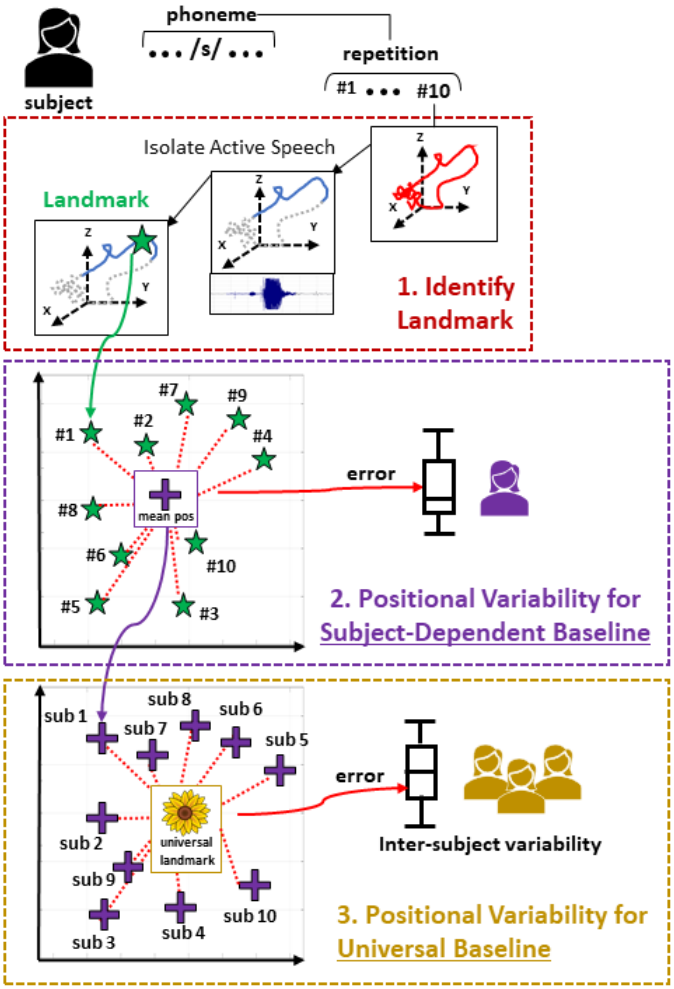

Fig. 4. Overview of our data analysis in which (1) the phoneme landmarks are identified from the tongue trajectories that were pre-processed by trimming out non-speech components, (2) positional variability are calculated in a subject-dependent case, and (3) the final landmarks for each phoneme and subject are combined together to generate a universal baseline.

the identification of the phoneme landmark. To better automate this step, a software was developed (Fig. 5) and is composed of the following parts: (1) a settings bar where the investigators can select a particular trajectory from the database, play the recorded voice, and set the start/stop time markers of the active speech period; (2) trajectory displayed as three time series (one per axis) with the positions recorded during active speech highlighted in color; (3) voice waveform with active speech highlighted in blue. The 3D position of a landmark is set by clicking on a target point on any of the time series, typically the longitude (Y) or height (Z), where the investigators estimated that it is representative of that phoneme (the other 2 values are automatically set based on the time stamp of the selected point). To further assist the investigators, a candidate landmark is automatically selected by an algorithm that implements the following steps for each axis: generate a histogram of the positions, and then select the bin with the highest distribution. This algorithm is only used to provide an initial estimation, but the investigators are tasked to set the final landmark by selecting this candidate or manually mark

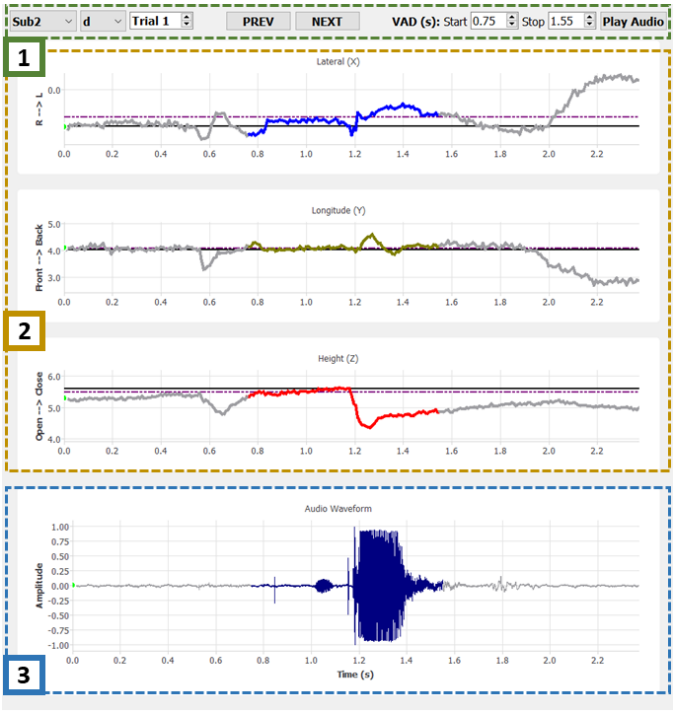

Fig. 5. User interface of the phoneme landmark identification program that displays the tongue trajectory as three time series, one per axis, and its associated voice waveform.

another position.

2) Positional Variability for Subject-Dependent Baseline: The objective of this step is to measure the positional variability of the phoneme landmarks for each subject individually to evaluate the ability of our system to generate a subject-dependant baseline. First, a subject's final phoneme landmark is set as the mean position of all the 10 repetition landmarks for that phoneme. Then, for each phoneme's repetition, a tracking error is defined as the Euclidean distance between that repetition and its final landmark. Finally, a positional variability is set per subject and by combining all the tracking errors together.

3) Positional Variability for Universal Baseline: The end goal of the overall research is to generate a baseline of phoneme landmarks that can be used for anyone, i.e. a baseline that is subject independent (referred here as universal). Since a final phoneme landmark was identified for each subject in the previous step, a universal phoneme landmark can be defined as the mean position among all the subjects' landmarks for a given phoneme. Similar to the previous step, an error can be measured as the Euclidean distance between each subject's and its universal phoneme landmark. An inter-subject variability is then created by combining these errors.

III. Results

Fig. 6 shows a summarized view of all tracking errors (top right) along with more detailed views of the errors split by subject (top left) and by phonemes (bottom). Overall, 75% of the phoneme landmarks identified from the tongue motion are within 5.8 mm of their estimated true position, while the median tracking error is 3.9 mm. Compared to the tracking errors reported in our in-lab test setup [15], these results are roughly three times higher than the testing set (median: 1.4, Q3: 1.8) but only 2x higher than the validation set (median: 2.3, Q3: 3.1). The highest errors are found for subject #7 with a Q3 of 7.4 mm, and the lowest for subject #2 with a Q3 of 4.2 mm. Interestingly, there is no demographically distinctive features that were found between these two subjects: both were born and raised in state of Georgia, are the same age (22 y.o), same ethnicity, and have similar accent (slightly southern). Regarding the split by phonemes, the variability is more consistent with slightly higher errors for the /r/ sounds and lower values reported for the consonants.

For the universal baseline, the inter-subject variability is shown in Fig. 7 and split across phonemes. Overall, the errors are much higher (median: 55 mm, Q3: 65 mm) and less consistent (interquartile range: >20 mm). These results are an order of magnitude (>10x) higher than the ones reported for the subject-dependent case and show that the generation of a universal baseline of phoneme landmarks is not feasible with the current version of our tongue tracking hardware and software. A discussion of the potential reasons for such a limitation of our system is provided in the following section.

IV. Discussion

As far as we know, this work is the first attempt to produce a universal baseline of phoneme landmarks. Although the generation of a universal baseline is not the objective of our current study but the ultimate goal of the overall research, it allows us to better identify the limitations of our current system and facilitates the creation of a roadmap for our future work. For instance, because the subject-dependent errors are much lower, a poor alignment of the trajectories between subjects could be the main reason for such high errors. Indeed, since the head of each subject is placed at a different position and orientation in the headset, the tongue trajectories for each subject were projected into a common frame of reference for inter-subject comparison. This projection of coordinates relied on estimating the position of five specific points in the oral cavity using a method similar as that in [19]. This procedure was not tested with our system, and thus can introduce significant errors due to alignment rather than tracking. Although unlikely as significant as alignment errors, additional errors are introduced by the fact that there is no method currently available to ensure that the tracer is placed at the same position on the tongue between subjects. Also, as reported in [19], the morphology and dimension of the oral cavity have an impact on the placement of phonemes. In future studies, we will verify that a reduction of these effects would decrease the differences in errors between the subject-dependent and universal cases.

For the subject-dependent baseline, the shortcomings of the universal case are generally not applicable since, as long as the data collection occurs in a single session, there is no need for a projection of coordinates and to account for differences in tracer placement and mouth morphology. However, errors unrelated to tracking can still occur. For instance, the process of selecting a landmark by a human, either by approving the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TBME.2020.3023284, IEEE Transactions on Biomedical Engineering

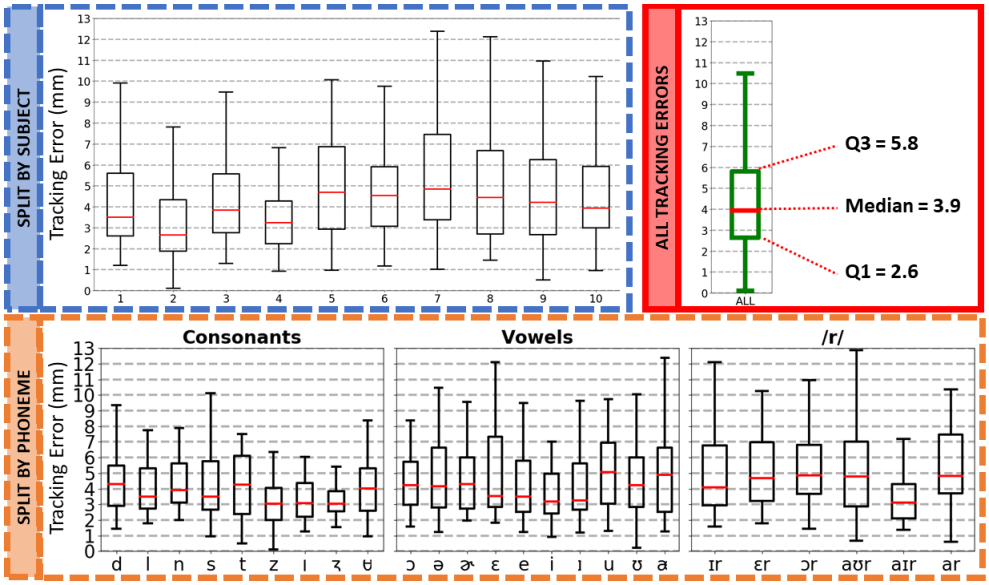

Fig. 6. Box plots of the tracking errors. The top-left plot shows the errors split by subject, the bottom plot by phoneme, and the top-right plot shows the final result with the tracking errors calculated across all landmarks.

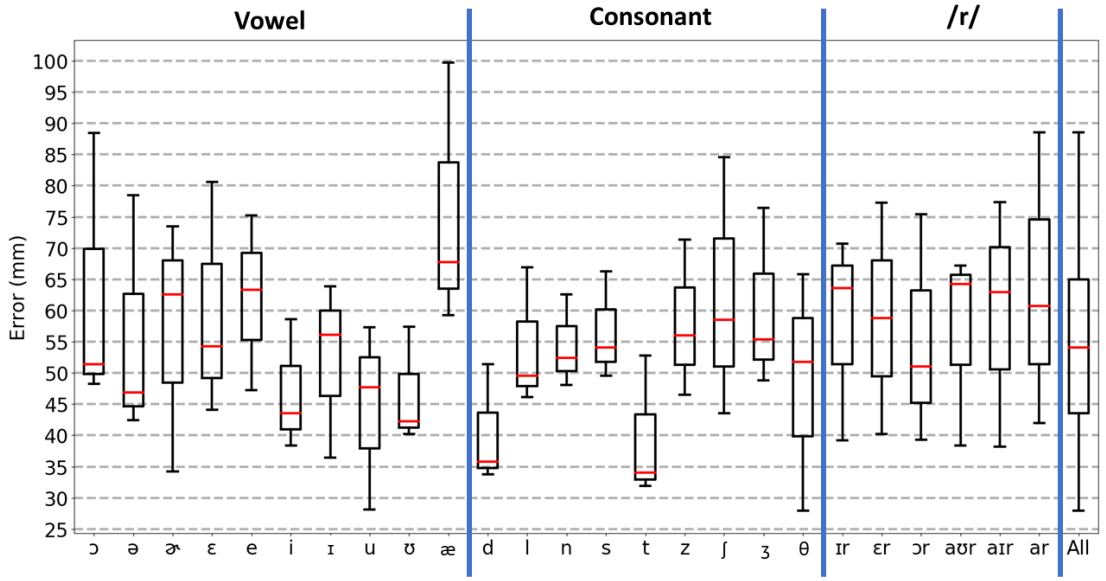

Fig. 7. Inter-subject variability of the positional errors for the phoneme landmarks, with the right-most box plot being the variability with all phonemes combined.

automated selection or manual identification, can add errors since this method is prone to subjectivity. Regarding the errors split by phonemes, the consonants have generally lower errors and this could be explained by the fact that the placement of the tongue's blade has a more important impact on the production of these phonemes. Indeed, the tongue's blade is in contact with the palate and/or front teeth, which increases the proprioception of the position of the tongue, and thus improves our ability to accurately place the tongue in a same location. In addition, higher errors in the /r/ phonemes may be attributed to different tongue configurations that can produce that phoneme [21]. These differences can occur both across and within speakers [22].

It is challenging to make an insightful analysis of the results because the science in speech sound production is still in its infancy with only few studies that have attempted to record tongue motion during speech production including, but not limited to, the studies in [7], [12], [18], [23], [24]. There are many reasons for the lack of research in that field such as the fact that tracking the tongue during speech and without impeding its natural motion is difficult to achieve, and because of the inherent complexity in articulation, motor control, voicing, among other mechanisms involved in speech production. For instance, the majority of the phonemes were produced in isolation while some consonants were followed by a vowel (C-V sequence). We found that it was more challenging to identify a characteristic landmark when a phoneme was produced in isolation and it might be more more difficult for the subjects

to place their tongue in a consistent location when producing the phonemes in that manner instead of a C-V-C or V-C-V sequence as done in other studies [19]. Additionally, it was observed that the subjects were not consistent between each other in how they uttered some phonemes even though our data collection was designed to reduce such variability. Indeed, a reference sound was played when a new phoneme was displayed, a word was shown alongside the target phoneme to provide a context for its pronunciation, and the subjects were CMSD students in the same program that recently passed a phonetic class where they learned how to properly produce these phonemes.

Before attempting to conduct a follow-up human study of clinical value, solutions need to be found to some important technical challenges. First, we must find a method to either place the tracer at the same position and in a consistent manner, or to account for this offset and post-process the trajectories. Secondly, the dimension of the oral cavity and the tongue must be taken into account when comparing the landmark positions between subjects as indicated in [19]. Third, the current headset version is stationary and thus forces the user to remain still for an optimal tracking accuracy because our system cannot remove magnetic disturbances, mainly due to interference from the Earth's magnetic field, when the headset is moving. However, natural body/head motion cannot be fully restricted and thus adds motion artifacts in the recorded trajectories that is not from the tongue. Future plans include the development of an algorithm to dynamically remove the magnetic disturbances, similar to the method described in [25], which will allow the user to freely move while wearing the headset and ensuring a satisfactory tracking accuracy. Finally, this study evaluates the repeatability in tracking but not its precision since the actual value of the tracer's positions are unknown. EMA could serve as a ground truth but the two systems cannot be used at the same time because they will interfere with each other. A subject could be asked to utter the same sequence of phonemes in one system at a time, but this will add some unknown, though likely not significant, variability due to the fact that the landmarks positions will be compared from different trajectories. Another possibility would be to place the tracer at known positions in the oral cavity, such as teeth and special oral landmarks [18], [19], but this would provide only a few data points for our evaluation of tracking precision. The combination of these methods might, at least, provide some indication of the tracking precision during actual speech.

Among the main limitations of our system is the fact that only one tracer can be tracked at a time. Undoubtedly, tracking multiple tracers simultaneously will provide more articulatory information about the tongue. A multi-magnet localization is an ongoing research in this field [26] but the tracking accuracy will likely decrease as compared to a single magnet due to the increased complexity of solving the nonlinear optimization in traditional localization methods or the exponential increase in the size of the training dataset with each added magnet in our method based on a neural network. Nonetheless, until a multi-tracer magnetic localization is available, tracking one tracer remains useful since studies such as [24], [27] show that tracking the tongue tip, along with upper and lower lips that can be tracked using more conventional methods (e.g. reflective trackers, computer vision), allows for speech to be recognized with high accuracy (>90% word recognition [27]). Although some phonemes (e.g. vowels) might likely be challenging to uniquely identify using one flesh point, the tracer can be placed in a different location on the tongue for each phoneme to generate the most useful visualization of tongue placement. Furthermore, our system is designed to be used in conjunction with other tongue placement tools, such as tongue tips or bite blocks, to enhance the range of treatment tools available to SLPs during speech therapy.

V. Conclusion

The creation of a baseline of phoneme landmarks was carried out in a human study in which 10 CMSD students were asked to utter 25 phonemes comprised of consonants, vowels, and variation of /r/ sounds. These phonemes were selected because of the importance of tongue placement for their production. A new wireless tracking system captured the motion of the tongue while a subject was uttering 10 repetitions for each phoneme. The positional variability of the phoneme landmarks was first calculated in the subject-dependent case, and the final landmarks for each subject were used to evaluate the feasibility of a universal baseline. We found that the inter-subject positional variability is sufficiently high (median: 55 mm, Q3: 65 mm) to conclude that a universal baseline cannot be generated by our current system. However, the positional variability in the subject-dependent case is an order of magnitude lower (median: 3.9 mm, Q3: 5.8 mm) which shows that the differences are unlikely due to a poor tracking accuracy of our system but mostly from technical challenges inherent to comparing tongue trajectories between subjects in which normalization and alignment have a stronger impact. Furthermore, the tracking errors reported for the subject-dependent baseline are effectively upper bounds since they include natural tongue placement variability. Regardless, in future work, the tracking error of our system must be reduced to provide positional variability of phoneme landmarks that are comparable to EMA [18] while enabling our system to be fully wearable. Once satisfactory tracking accuracy will be reached, our tongue tracking system will enable researchers in speech science to collect valuable information in a way that minimally impacts natural speech thanks to its wireless tracking method. Also, the system is affordable because it is composed of mass-produced components which will enable more researchers to access this tool and collect data to increase the body of knowledge about the influence of the tongue in speech production. More importantly, the overall objective is to enable our system to be used in speech therapy where it can help millions of people affected by speech sound disorders to achieve increased intelligibility, and thus, improve their quality of life.

Acknowledgment

The authors would like to thank the CMSD students at the University of Georgia for their participation in this study.

REFERENCES

[1] J. R. Duffy, *Motor Speech disorders: Substrates, differential diagnosis, and management*. Elsevier Health Sciences, 2013, p. 512.

[2] NIDCD, "Statistics on voice, speech, and language," National Institute on Deafness and Other Communication Disorders, Report, 2016. [Online]. Available: https://www.nidcd.nih.gov/health/statistics/statistics-voice-speech-and-language

[3] W. R. Zemlin, *Speech and Hearing Science: Anatomy and Physiology*, 4th ed. Pearson, 1998, p. 610.

[4] P. Marshalla, "Horns, whistles, bite blocks, and straws: a review of tools/objects used in articulation therapy by van riper and other traditional therapists." *International Journal of Orofacial Myology*, vol. 37, pp. 69–96, 2011.

[5] T. M. Byun and E. R. Hitchcock, "Investigating the use of traditional and spectral biofeedback approaches to intervention for /r/ misarticulation," *American Journal of Speech-Language Pathology*, vol. 21, no. 3, pp. 207–221, 2012.

[6] W. Katz, T. F. Campbell, J. Wang, E. Farrar, J. C. Eubanks, A. Balasubramanian, B. Prabhakaran, and R. Rennaker, "Opti-speech: A real-time, 3d visual feedback system for speech training," in *Fifteenth Annual Conference of the International Speech Communication Association*, 2014, pp. 1174–1178.

[7] W. F. Katz and S. Mehta, "Visual feedback of tongue movement for novel speech sound learning," *Frontiers in human neuroscience*, vol. 9, p. 612, 2015.

[8] P. Schoenle, K. Grbe, P. Wenig, J. Hhne, J. Schrader, and B. Conrad, "Electromagnetic articulography: Use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract," *Brain and Language*, vol. 31, no. 1, pp. 26–35, 1987.

[9] T. Kaburagi, K. Wakamiya, and M. Honda, "Three-dimensional electromagnetic articulography: A measurement principle," *Journal of the Acoustical Society of America*, vol. 118, no. 1, pp. 428–443, 2005.

[10] S. Kelly, A. Main, G. Manley, and C. McLean, "Electropalatography and the linguagraph system," *Medical Engineering & Physics*, vol. 22, no. 1, pp. 47–58, 2000.

[11] F. Gibbon and A. Lee, "Electropalatography for older children and adults with residual speech errors," *Semin Speech Lang*, vol. 36, no. 04, pp. 271–282, 2015.

[12] W. F. Katz, S. V. Bharadwaj, and M. P. Stettler, "Influences of electromagnetic articulography sensors on speech produced by healthy adults and individuals with aphasia and apraxia," *Journal of Speech, Language, and Hearing Research*, vol. 49, no. 3, pp. 645–659, 2006.

[13] N. Sebkhi, D. Desai, M. Islam, J. Lu, K. Wilson, and M. Ghovanloo, "Multimodal speech capture system for speech rehabilitation and learning," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 11, pp. 2639–2649, 2017.

[14] N. Sebkhi, Y. Yunusova, and M. Ghovanloo, "Towards phoneme landmarks identification for american-english using a multimodal speech capture system," in *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, Oct 2018, pp. 1–4.

[15] N. Sebkhi, N. Sahadat, S. Hersek, A. Bhavsar, S. Siahpoushan, M. Ghovanloo, and O. Inan, "A deep neural network-based permanent magnet localization for tongue tracking," *IEEE Sensors Journal*, pp. 1–1, 2019.

[16] Y. Yunusova, J. R. Green, and A. Mefferd, "Accuracy assessment for ag500, electromagnetic articulograph," *Journal of Speech, Language, and Hearing Research*, vol. 52, no. 2, pp. 547–555, 2009.

[17] J. J. Berry, "Accuracy of the ndi wave speech research system," *Journal of Speech, Language & Hearing Research*, vol. 54, no. 5, pp. 1295–1301, 2011.

[18] Y. Yunusova, J. S. Rosenthal, K. Rudy, M. Baljko, and J. Daskalogiannakis, "Positional targets for lingual consonants defined using electromagnetic articulography," *Journal of the Acoustical Society of America*, vol. 132, no. 2, pp. 1027–1038, 2012.

[19] K. Rudy and Y. Yunusova, "The effect of anatomic factors on tongue position variability during consonants," *Journal of Speech, Language, and Hearing Research*, vol. 56, no. 1, pp. 137–149, Feb 2013.

[20] J. Reilly and J. L. Fisher, "Sherlock holmes and the strange case of the missing attribution: A historical note on the grandfather passage," *Journal of Speech, Language, and Hearing Research*, 2012.

[21] S. Boyce, "The articulatory phonetics of /r/ for residual speech errors," *Seminars in speech and language*, vol. 36, pp. 257–270, 2015.

[22] C. Espy-Wilson, S. Boyce, M. Jackson, S. Narayanan, and A. Alwan, "Acoustic modeling of american english /r/," *The Journal of the Acoustical Society of America*, vol. 108, pp. 343–56, 2000.

[23] W. F. Katz and M. R. McNeil, "Studies of articulatory feedback treatment for apraxia of speech based on electromagnetic articulography," *Perspectives on Neurophysiology and Neurogenic Speech and Language Disorders*, vol. 20, no. 3, pp. 73–80, 2010.

[24] J. Wang, A. Samal, P. Rong, and J. R. Green, "An optimal set of flesh points on tongue and lips for speech-movement classification," *Journal of Speech, Language, and Hearing Research*, vol. 59, no. 1, pp. 15–26, Feb 2016.

[25] M. N. Sahadat, N. Sebkhi, D. Anderson, and M. Ghovanloo, "Optimization of tongue gesture processing algorithm for standalone multimodal tongue drive system," *IEEE Sensors J.*, vol. 19, no. 7, pp. 2704–2712, 2019.

[26] S. Song, C. Hu, and M. Q.-H. Meng, "Multiple objects positioning and identification method based on magnetic localization system," *IEEE Transactions on Magnetics*, vol. 52, no. 10, pp. 1–4, 2016.

[27] B. Cao, B. Tsang, and J. Wang, "Comparing the performance of individual articulatory flesh points for articulation-to-speech synthesis," in *Proceedings of the 19th International Congress of Phonetic Sciences*, 08 2019, Conference Proceedings, pp. 1–5.