

# Evaluation of Multimodal Tongue Drive System by People with Tetraplegia for Computer Access

Md Nazmus Sahadat<sup>1</sup>, Nordine Sebkhil<sup>1</sup>, Arpan Bhavsar<sup>1</sup>, Erica Walling<sup>2</sup>, Michelle Hoefnagel<sup>2</sup>, Fulcher Chris<sup>2</sup>, Fanpeng Kong<sup>1</sup>, Geneva Tonuzi<sup>2</sup>, Osborne Raine<sup>2</sup>, Maysam Ghovanloo, *Fellow, IEEE*, and David Anderson<sup>1</sup>, *Senior Member, IEEE*

**Abstract**—The multimodal Tongue Drive System (mTDS) is an assistive technology for people with tetraplegia that provides an alternative method to interact with a computer by combining tongue control, head gesture, and speech. This multimodality is designed to facilitate the completion of complex computer tasks (e.g. drag-and-drop) that cannot be easily performed by existing uni-modal assistive technologies. Previous studies with able-bodied participants showed promising performance of the mTDS on complex tasks when compared to other input methods such as keyboard and mouse. In this three-session pilot study, the primary objective is to show the feasibility of using mTDS to facilitate human-computer interactions by asking fourteen participants with tetraplegia to complete five computer access tasks with increased level of complexity: maze navigation, center-out tapping, playing bubble shooter and peg solitaire, and sending an email. Speed and accuracy are quantified by key metrics that are found to be generally increasing from the first to third session, indicating the potential existence of a learning phase that could result in improved performance over time.

**Index Terms**—Assistive technology, human computer interaction, machine learning, spinal cord injury, tetraplegia, tongue drive system.

## I. INTRODUCTION

EVERY year, an estimated 500,000 people worldwide have accidents that result in a spinal cord injury (SCI) which has many negative consequences to their overall quality of life including lower rates of school enrollment or employment, and higher financial burden on their household besides the societal cost [1]. In the United States, SCI is the leading cause of tetraplegia, i.e. paralysis of the four limbs, resulting in direct healthcare cost estimated at around \$1M during the first year and \$100k per subsequent years [2]. Additionally, more than 56% of people with tetraplegia are between the ages of 16-30 years old and the majority (76%) lives in a household with annual income below \$50k [3].

Because people with high-level tetraplegia cannot use their four limbs, this population is heavily relying on caregivers for their activities of daily living (ADL). Such activities include using a keyboard and a mouse to access a computer, using a finger to control a smartphone, or even driving their own

This work was supported in part by the NSF awards CBET-0828882 and IIS-0803184. Corresponding author: sebnor31@gatech.edu

<sup>1</sup>School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA.

<sup>2</sup>Brooks Rehabilitation, Jacksonville, FL, USA.

This work has been submitted to the IEEE for possible publication. Copyright may be transferred without notice, after which this version may no longer be accessible.

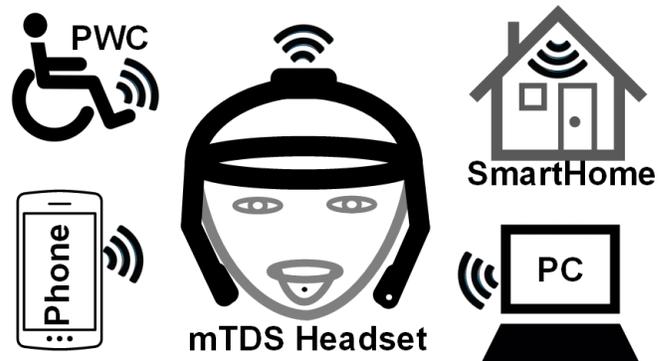


Fig. 1. The standalone multimodal tongue drive system (mTDS) converts tongue motion, head gesture, and speech, into commands that are transmitted wirelessly to control many devices such as computers, smartphones, powered wheelchair (PWC), and smart-home devices [15]

powered wheelchair. In order to regain a degree of autonomy, assistive technologies (ATs) have been developed to leverage their remaining abilities for human-computer interaction such as tracking eye movements [4], muscle activity [5], [6], speech [7], [8], and brain-computer interfaces [9], [10]. The Tongue Drive System (TDS) is an assistive technology that can track tongue movements to issue commands to control devices. The TDS has many advantages over the commercially-available systems such as performing optimally in noisy environment, not being prone to muscle fatigue, and providing more privacy since the tongue is hidden inside the mouth [11]–[14].

The majority of ATs used by people with tetraplegia leverages a single modality for human-computer interaction. Although a unimodal AT is simpler to use, a multimodal system enables completion of more complex tasks. For instance, switches (e.g. push-button) are more suitable to issue discrete commands but they are not effective for proportional control such as moving a mouse cursor. Thus, the efficacy of an AT can be significantly improved by leveraging multiple remaining abilities. The multimodal Tongue Drive System (mTDS) is an example of such AT in which tongue movements are translated into discrete commands, head movement into proportional input, and speech is used for typing [15]–[18]. By combining these modalities, complex tasks can be performed with more ease and, as illustrated in Fig. 1, this enables the mTDS to control a wider range of devices in the user's surrounding, such as their powered wheelchair, smartphone, smart-home appliances, a computer, among many others.

Arguably, considering today’s lifestyle and the nature of most vocations, one of the most critical device interactions that a person with tetraplegia would need is computer access. Thanks to the multimodal capabilities of the mTDS, complex computer interactions can be performed, such as drag and drop, scrolling, and selecting multiple icons. In previous studies [16], [18], the mTDS has shown promising performance on complex computer tasks with 15 able-bodied participants being able to send emails at about twice the time compared to a standard keyboard-mouse combination [18]. Furthermore, the majority of participants were using the mTDS for the first time and were given only few attempts to perform these tasks in a single data collection session.

Since these previous studies involved able-bodied participants, their performance in completing their tasks with mTDS could be contrasted with a ground-truth in the form of the keyboard and mouse. In this pilot study, our objective is to evaluate the ability of individuals with tetraplegia, rather than able-bodied, to complete a set of computer access tasks using the mTDS. There is not one universal input method used by all people with tetraplegia that can serve as a baseline of performance. Therefore, this study is primarily interested in showing the feasibility of using mTDS as an assistive technology for human-computer interactions rather than a comparison with existing controllers.

Because there is a large variety of computer interactions, our focus in this study is on the control of the mouse cursor. The proportional modality of the mTDS with head motion has been used to control the position of the cursor/pointer on the screen, and one out of the seven discrete commands that can be issued by the tongue modality has emulated a mouse click. Four tasks was designed to evaluate the speed and accuracy of the user to control the mouse using these aforementioned modalities: navigating a maze, center-out tapping, playing bubble shooter and peg solitaire. Additionally, an email sending task will serve as an instance of a task that is not only representative of a real-world computer access task but also requires all modalities that the mTDS can provide, including speech for typing.

The rest of the paper is organized as follows: section II will provide a system overview of the mTDS and more information about the protocol for data collection, section III will provide more details on the computer access tasks along with their key performance metrics, section IV will show a summary of the results followed by a discussion in section V and a conclusion in section VI.

## II. METHODS

### A. mTDS: System Overview

The mTDS utilizes three remaining abilities for human-computer interaction that are typically retained by people with tetraplegia: speech, tongue movements, and partial head motion. These modalities are captured by sensors embedded in the mTDS headset (Fig. 2) and converted into commands specific to a target device [19]. In this pilot study, the application of interest is computer access and thus the commands are designed to emulate the mouse capability to move a cursor and perform clicks as well as typing using a keyboard. Two

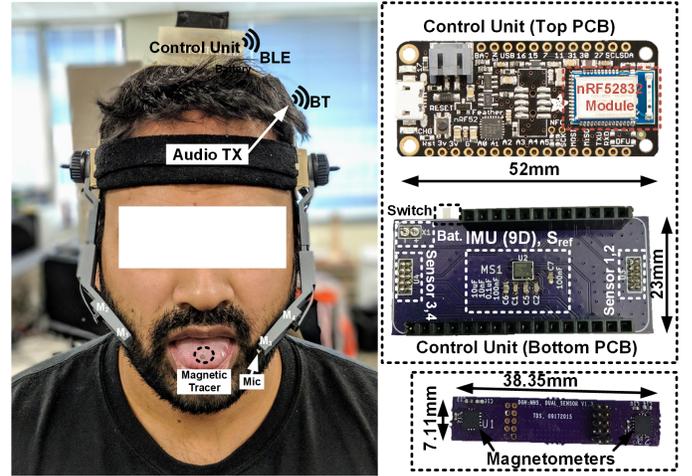


Fig. 2. mTDS headset composed of four magnetometers to capture the tongue tracer’s magnetic field, a microphone to record speech, and a control unit to measure head motion and transmit processed data wirelessly using bluetooth [15].

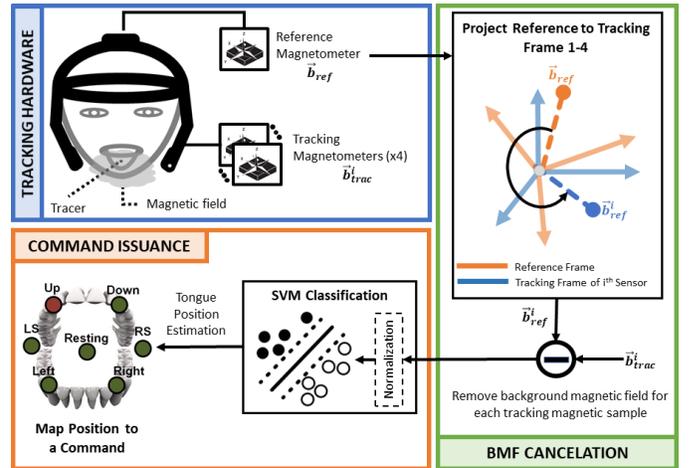


Fig. 3. Overview of the data processing that issues tongue commands from the magnetic measurements of the mTDS tracking hardware.

user-defined tongue positions in the mouth (e.g. touching the left and right cheek) are mapped to the left and right mouse clicks, head pitch and roll are converted into vertical and horizontal mouse cursor movements, and speech recognition for typing [16]. As shown in Fig. 3, the tongue position in the mouth is estimated from a small disk-shaped magnetic tracer ( $\phi$  3.2 mm x 1.6 mm) manufactured by K&J Magnetics (Jamison, PA, USA) that is attached to the tongue, about 1 cm posterior to the tip, using a cyanoacrylic adhesive (PeriAcryl, GluStitch Inc., Canada). When the tracer moves with tongue movements, it changes the magnetic field inside and around the mouth, which are measured at a rate of 100 Hz by four 3-axis tracking magnetometers embedded in the wearable headset. To isolate the magnetic field generated by the tracer, the background magnetic field (BMF) is measured by a reference magnetometer that is placed on top of the headset to be as far from the tracer as possible. For each tracking magnetometer, the BMF is first projected from its original reference frame to the frame of that magnetometer,

and then subtracted from the raw measurements. The BMF-canceled magnetic measurements are then normalized before being fed into a support vector machine (SVM) classifier that was trained on seven user-defined tongue positions. Each tongue position is mapped to a specific command for the target application. More details about the data processing for tongue-generated commands can be found in [15]. An inertial measure unit (IMU) is embedded in the control unit mounted on the top of the headset which captures head movements using an accelerometer and a gyroscope. A sensor fusion algorithm was developed to estimate the pitch and roll angles of the head as explained in more details in [19]. Speech is captured using a microphone attached to one of the headset's arms and audio signals are transmitted to a computer to perform speech recognition using Dragon Naturally Speaking (Nuance Communications Inc., Burlington, MA, USA). All data are transmitted wirelessly to a computer using Bluetooth communication for storage and post-processing.

### B. Protocol

Fourteen participants were recruited for this pilot study at Brooks Rehabilitation (Jacksonville, FL, USA) and approved by the Schulman Institutional Review Board. The inclusion criteria were being diagnosed with tetraplegia and no prior experience with the mTDS. All participants were males and within the age range of 20-62 years old. For each participant, the study was split into three sessions: first two sessions were conducted within a week, followed by the third session scheduled 7-10 days after the second session in order to assess learning retention of using the mTDS. In the first session, an instructional video was played to introduce the participant to the mTDS, and specifically how to use its various modalities and the tasks that must be completed. Each session was divided into three parts and limited to 3 hours to reduce the impact of fatigue to the study outcome. A new and sterilized magnetic tracer was attached to the subject's tongue at each session and at ~1 cm from the tip. Although the tracer attachment in this

study is temporary, it is nonetheless placed at the same location at which a tongue piercing will be found for long term use [20]. The participant is then asked to place the tracer at seven different locations in the mouth, for a duration of 25 seconds per location and a total of three repetitions, in order to train a machine learning classifier that can recognize the desired command from the magnetic field signature generated by the tracer. To assess the accuracy of the classifier, the participant is asked to place the tongue at these seven positions in a random manner and validate that the recognized commands are correct. Once this setup is successfully completed, the participant is then asked to perform a series of computer access tasks (Fig. 4) as explained in the following sections.

## III. COMPUTER ACCESS TASKS

In this section, each computer access task will be introduced and they are sorted by the number of modalities required for completion. Maze navigation (III-A) is an example of a unimodal task that only requires head motion, while two modalities (tongue and head motion) are needed for center-out tapping (III-B), bubble shooter (III-C) and peg solitaire (III-D). Lastly, sending emails (III-E) is not only an example of a realistic and useful computer access task but also of a complex task that requires all modalities: tongue/head motion and speech. For each task, an evaluation of the ability of the participants to complete that task using the mTDS has been defined as the average across all participants' performance for key metrics. For all the equations shown in this section,  $n$  is the number of participants and  $t$  is the number of trials.

### A. Maze Navigation

This task evaluates the efficacy of controlling the movement of a cursor in a space-constrained setting and using head motion only (Fig. 5). It also has similarities to wheelchair navigation. The participant was asked to move the mouse cursor inside a pre-defined track from a starting location to an endpoint as rapidly and accurately as possible. To increase the difficulty of this task, the width of the track decreases when approaching the endpoint from 50 to 20 pixels. Each maze was divided into 21 segments with the transition from one segment to another requiring a change in direction by  $45^\circ$ ,  $90^\circ$ ,  $135^\circ$ ,  $225^\circ$  or  $315^\circ$ . The duration of this task was limited to 2 minutes per round, and the layout of the maze was randomly selected from 8 pre-defined patterns with a similar level of difficulty (length:  $2262 \pm 48$  pixels) to prevent adaptation bias [16].

The performance metric selected for this task is the throughput,  $TPm$  (in bits/s), that combines both the speed and accuracy of the input device during cursor navigation in this space-constrained path [21] and is defined in equation (1),

$$TPm_{avg} = \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{t} \sum_{j=1}^t \left( \frac{ID_p P_{in} (1 - P_{out})}{CT} \right)_{ij} \right) \quad (1)$$

where  $P_{in}$  and  $P_{out}$  are the percentages of distance that the cursor has traversed inside and outside of the maze, respectively [21]. Maze completion time,  $CT$ , is the time



Fig. 4. Illustration of a participant using the mTDS for a computer access task.

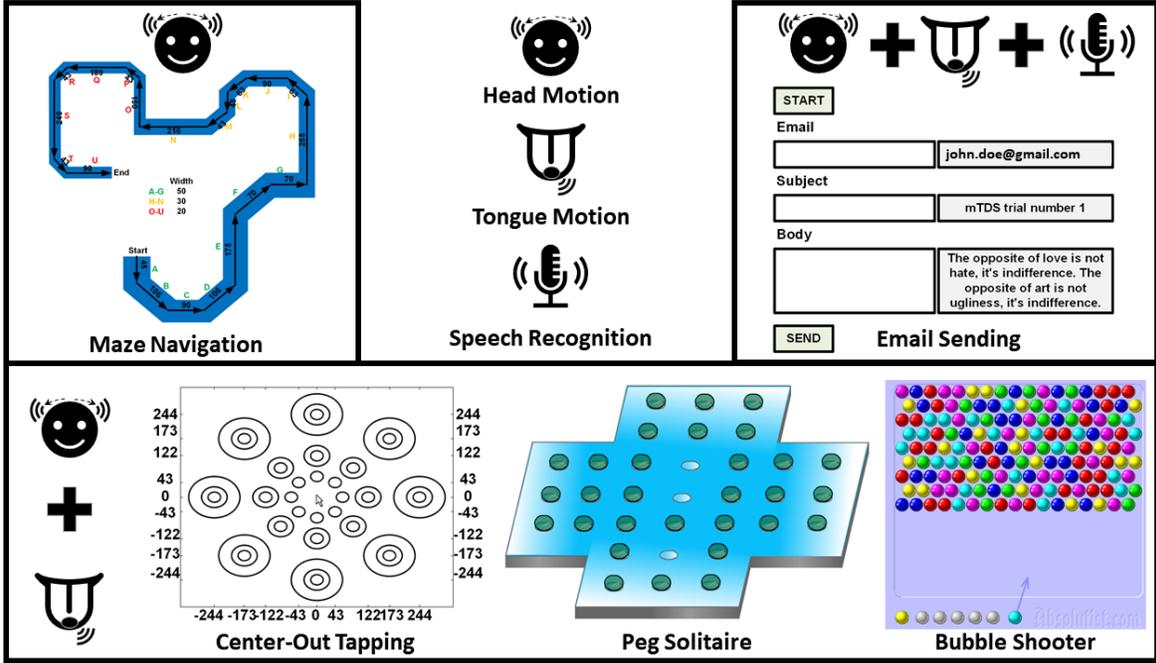


Fig. 5. Five tasks performed by the participants with varying number of modalities required to be completed: a unimodal task with head motion only (maze navigation), bimodal tasks with head and tongue motion (center-out tapping, peg solitaire, and bubble shooter), and a trimodal task that adds speech recognition (email sending). Maze navigation consists of following a path with the mouse cursor which is controlled by head motion. Center-out tapping consists of placing the mouse cursor inside a target with varying dimension and position. Peg solitaire uses both head and tongue modalities simultaneously to perform drag-and-drops of pegs in empty slots. Bubble shooter relies on head motion to set a direction of an arrow and tongue modality to shoot the bubble in that direction. Sending an email requires typing an address, subject, and body of the email using speech, navigating the cursor to each text field and the Send button with head motion, and clicking the button using the tongue modality.

required for a participant to navigate from start to endpoint, and  $t$  is the total number of trials. Path index of difficulty,  $ID_P$  in *bits*, is calculated using Shannon's formula (2),

$$ID_P = \sum_{i=1}^s \log_2 \left( \frac{L_i}{W_i} + 1 \right) \quad (2)$$

where  $s$  is the number of maze segments,  $L$  is the length and  $W$  is the width of a segment of the maze.

### B. Center-out Tapping

Center-out tapping task evaluates the performance of pointing devices using Fitts law [22]. As illustrated in Fig. 5, participants were instructed to move the cursor from the center of the screen to a round target that is placed randomly within a pre-defined set of diameters (30, 60 and 122 pixels) and distances (61, 122, 244 pixels). Only one target is shown at a time and the participant does not know the sequence of targets. The movement of the pointer, i.e. mouse cursor in this study, is controlled using the head motion (pitch and roll) and a target selection via mouse click is emulated by touching the left cheek with the tongue. A new target appeared upon selecting the previous one for a total of 48 random targets per round. More details about the performance metrics are provided below.

1) *Throughput (TP)*: The throughput measures the information transfer rates from a pointing device to a computer in bits/s. It is calculated using the ratio between the index of

difficulty ( $ID$ ) and the average movement time ( $MT$ ) [22], [23] as shown in (3),

$$TP_{avg} = \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{m} \sum_{j=1}^m \frac{ID_{ij}}{MT_{ij}} \right) \quad (3)$$

where  $m$  is the total number of targets, and  $ID$  is calculated using Shannon's formula as shown below,

$$ID = \log_2 \left( \frac{D}{W} + 1 \right) \quad (4)$$

where  $D$  is the distance from the center of the screen to the center of the target and  $W$  is the diameter of the circular target. Higher  $ID$  targets are more difficult to reach because the targets are smaller and farther, and thus require more information to be transferred.

2) *Reaction Time (RT)*: This metrics represents the elapsed time (in seconds) between the display of a target and the point in time at which the participant starts moving the cursor towards it [20]. The reaction time, which also includes the system latency, is reported as the mean of mean value using equation (5),

$$RT_{avg} = \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{m} \sum_{j=1}^m RT_{ij} \right) \quad (5)$$

3) *Error Rate (ER)*: This error rate is the percentage of errors for the selection of a target defined by an  $Error^{select}$  parameter that is either '1' if the selection is performed outside of the target or '0' otherwise.  $ER_{avg}(\%)$  is the ratio of the sum of error selections and the total number of selections calculated using equation (6) [20],

$$ER_{avg}(\%) = \frac{\sum_{j=1}^x Error_j^{select}}{\sum_{i=1}^y selection_i} \times 100 \quad (6)$$

where  $x$  is the total number of selections outside of the target and  $y$  is the total number of selections for all participants and throughout the 3 sessions.

4) *Path Efficiency (PE)*: Equation (7) shows the path efficiency as defined by the ratio between the actual cursor's path traversed by the participant ( $Path^{sub}$ ) and the most optimal path measured as the euclidean distance from the starting position (i.e. center of the screen) to the target center ( $Path^{opt}$ ) [20],

$$PE_{avg}(\%) = \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{m} \sum_{j=1}^m \frac{Path_j^{opt}}{Path_j^{sub}} \times 100 \right) \quad (7)$$

### C. Bubble Shooter

This task evaluates the sequential use of the head and tongue motion modalities. In this task (Fig. 5), the direction of the arrow is set according to the head's roll angle and shooting the bubble is performed by touching the left cheek with the tongue as a substitute for mouse left click. A time limit of 2 minutes per round was set and the objective was to place a bubble next to three or more bubbles of the same color, after which this set of color-like bubbles popped out and the score was increased. A Shooting Score ( $SS$ ) was calculated as follows,

$$SS_{avg} = \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{t} \sum_{j=1}^t score_{ij} \right) \quad (8)$$

where  $score$  is the score that a participant achieves at the end of a round.

### D. Peg Solitaire

This task is playing peg solitaire to evaluate the efficacy of using two modalities, tongue and head movements, to perform a drag-and-drop. The participant uses head motion to place the mouse cursor over a peg, selects a peg by touching the left cheek with the tongue, and drags the peg to another location using head motion. The objective is to precisely drop the dragged peg on top of an empty slot to consider this drag-and-drop action to be successful. Otherwise, the peg bounces back to its original slot and this drag-and-drop attempt is ignored. Similar to other tasks, the time was limited to 2 minutes per round, after which the number of successful drag-and-drop occurrences ( $DD$ ) are calculated as a performance metric,

$$DD_{avg} = \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{t} \sum_{j=1}^t DD_{ij} \right) \quad (9)$$

### E. Sending an Email

Sending an email is one of the most fundamental tasks that a user performs on a computer. Although simple to perform with a keyboard and mouse for an able-bodied user, it is more complex and somewhat cumbersome to achieve when using ATs because it requires multiple actions: cursor navigation, clicking, and text entry. In our study, participants were asked to compose an email which content was randomly selected among 60 sample texts with comparable length and complexity: 2-3 sentences comprised of  $24.8 \pm 4.7$  words. The time to complete this task was limited to 5 minutes, higher than the other tasks because of its complexity. A user interface was custom designed using LabVIEW (Fig. 5) to record the subject's actions and facilitate the estimation of the following metrics.

1) *Task Completion Time (TCT)*: This metric represents the time required by a participant to complete this task (in minutes) [16]. It is calculated using the following equation,

$$TCT_{avg} = \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{t} \sum_{j=1}^t TCT_{ij} \right) \quad (10)$$

where  $TCT$  is the completion time for a single trial.

2) *Text Entry Error Rate (TEER)*: This metric is calculated using the minimum string distance (MSD) as shown below,

$$TEER_{avg}(\%) = \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{t} \sum_{j=1}^t ErrorRate_{ij} \right) \quad (11)$$

where  $ErrorRate$  is defined as follows,

$$ErrorRate(\%) = \frac{MSD(A, B)}{\max(|A|, |B|)} \times 100, \quad (12)$$

where  $A$  and  $B$  represent the texts that were entered and targeted, respectively.  $MSD$  is calculated using the definition in [24], and  $\max(|A|, |B|)$  is the maximum number of characters between the two texts.

3) *Typing Speed (TS)*: The typing speed represents the number of words per minute (wpm) entered by the participant, and this metric includes delays originated from cursor navigation and clicking. Its equation is shown below,

$$TS_{avg} = \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{t} \sum_{j=1}^t \frac{\#words_{ij}}{TCT_{ij}} \right) \quad (13)$$

## IV. RESULTS

Table I provides a summary of the performance metrics calculated over the three sessions with all results shown as a mean value  $\pm$  standard deviation across the participants. A statistical analysis of the performance improvement is also provided in Table II as a one-tailed paired t-test to assess the learning ability of the participants.

TABLE I  
PERFORMANCE METRICS (*mean ± sem*) OVER THREE SESSIONS

Metrics	Session 1	Session 2	Session 3
Maze Navigation			
<i>TPm</i> (bps)	0.31 ± 0.03	0.42 ± 0.03	0.47 ± 0.04
Center-out Tapping			
<i>TP</i> (bps)	0.85 ± 0.01	1.04 ± 0.03	1.11 ± 0.02
<i>RT</i> (s)	0.94 ± 0.02	0.88 ± 0.02	0.79 ± 0.02
<i>ER</i> (%)	23.4 ± 14.2	29.5 ± 16.2	25.7 ± 19.1
<i>PE</i> (%)	62.6 ± 0.7	57.7 ± 0.8	60.5 ± 0.8
Bubble Shooting			
<i>SS</i>	766 ± 91	906 ± 71	1124 ± 127
Peg Solitaire			
<i>DD</i>	4.4 ± 0.37	6.5 ± 0.5	7.7 ± 0.8
Email Sending			
<i>TCT</i> (min)	0.87 ± 0.06	0.65 ± 0.02	0.64 ± 0.02
<i>TEER</i> (%)	32.4 ± 4.4	17.4 ± 2.6	31.6 ± 4.7
<i>TS</i> (wpm)	41.3 ± 2.7	55.5 ± 2.5	49.2 ± 3.1

TABLE II  
STATISTICAL ANALYSIS (ONE-TAILED T-TEST) : LEARNING ANALYSIS  
FROM FIRST TO LAST SESSION

	<i>t</i>	<i>t<sub>cr</sub></i>	<i>df</i>	<i>p</i>	<i>H<sub>0</sub></i>
Maze Navigation					
TPm (bps)	5.52	1.83	9	0.00002	rejected
Center-out Tapping					
TP (bps)	2.71	1.83	9	0.012	rejected
RT (s)	1.52	1.83	9	0.082	accepted
ER (%)	-0.07	1.83	9	0.53	accepted
PE (%)	-0.04	1.83	9	0.52	accepted
Bubble Shooting					
*SS	3.01	1.83	9	0.007	rejected
Peg Solitaire					
DD	3.54	1.86	8	0.0038	rejected
Email Sending					
TCT (min)	3.03	1.83	9	0.007	rejected
TEER (%)	-0.52	1.83	9	0.69	accepted
TS (wpm)	2.78	1.83	9	0.011	rejected

*t*: one sided t-value, *t<sub>cr</sub>*: critical t-value with  $\alpha = 0.05$ ,  
*df*: degree of freedom, *p*: p value, *H<sub>0</sub>*: Null hypothesis.  
\*statistically significant

#### A. Maze Navigation

The throughput (*TPm*) captures both speed and accuracy for this navigation task. A significant difference between the first (0.31 bps) and last session (0.47 bps) was observed which shows that the participants learned over the three sessions. This resulted in an improvement of 35%, 13%, and 52% from first to second, second to third, and first to third session, respectively. The variability remains low with standard deviation below 0.04 bps.

#### B. Center-out Tapping

A significant difference was observed for the throughput (*TP*) which improved from 0.85 bps to 1.11 bps (30.7%) from the first to the third session. Similarly to the maze navigation throughput, the increase from the second to the third session is less pronounced at 6%. The variability is also rather low with standard deviations below 0.03 bps.

The reaction time (*RT*) for the first, second, and third sessions were 0.94s, 0.88s, and 0.79s, which amounts to an improvement of 6%, 11%, and 16%, respectively. The results

are highly consistent with standard deviations representing less than 2% of the mean value.

Contrary to the aforementioned metrics, a different trend was observed for the error rate (*ER*) with percentages evaluated at 23.41%, 29.49%, and 25.72%. Therefore, the errors increased by 26% from the first to second session, but decreased by 13% from second to third session. This results in an overall increase by 10% from first to third session. However, high variability is reported for each session, which can be attributed to the fact that some participants attempted to decrease their movement time (i.e. time to move the cursor to a final position) while others preferred to be more accurate in selecting inside the target.

Although the improvement of path efficiency was not an objective of this task, it remains interesting to observe if added experience with the mTDS has any impact on how efficiently the cursor was moved to reach the targets. We can observe that the path efficiency (*PE*) remains consistently around 60% across all sessions and participants.

Overall, no significant differences were found for the metrics of this task except for the throughput.

#### C. Bubble Shooting

The overall shooting score (*SS*) represents both the speed and accuracy of this task. Its increase was statistically significant across the sessions (66, 905, and 1123) which translates to improvement of 18%, 24%, and 46% from first to second, second to third, and first to third session, respectively. The variability is moderate with standard deviations representing about 10% of the mean value, which could be explained by the difference in playing strategy between participants with some preferring to shoot bubbles at a faster pace while others preferred to spend more time thinking about their next actions.

#### D. Peg Solitaire

This task is also an example of both speed and accuracy, along with some cognitive load, needed to improve its associated metrics, which in this case is the number of drag-and-drop (*DD*) successfully performed by the participant. The increase of performance was statistically significant with 4.43, 6.53, and 7.7 drag-and-drops completed for the first, second, and third session, respectively. This translates to improvements of 47%, 18%, and 74% from first to second, second to third, and first to third session, respectively. Similarly to bubble shooting, the variability is moderate with standard deviations at about 10% of the mean values.

#### E. Email Sending

The task completion time (*TCT*) has been decreasing from 0.87 min, 0.65 min, to 0.64 min, for the first, second, and third session, respectively. This amounts to a statistically significant improvement of 26% from first to second session. Although there is no noticeable change in the average *TCT* from the second to third session, its consistency has nonetheless drastically improved with their standard deviations amounting to a third of the first session.

The text entry error rate ( $TEER$ ) represents the accuracy of this task and one of the metrics that only evaluates the performance of the speech recognition modality of the mTDS. There is no significant changes in the  $TEER$  between the first and third session with errors measured at around 32%, but interestingly, a major improvement was consistently reported for the second session with a  $TEER$  valued at 17%. This result might indicate that the learning phase between the first and second session increases the participant's proficiency with the speech recognition software, but this proficiency is not retained if a certain period of time has elapsed.

The typing speed ( $TS$ ) is the other metric that only evaluates speech recognition and it relates to the speed of its algorithm to generate words. With our system,  $TS$  was estimated at 41 wpm, 55 wpm, and 49 wpm during the first, second, and third session, respectively. Similarly to the text entry error rate, its improvement between the first and third session is not significant (<20%), but an increased in typing speed is observed for the second session with a 35% improvement from the first to the second session.

## V. DISCUSSION

Overall, the trend of the performance metrics is consistent increase throughout the three sessions of this pilot study, showing evidence that the participants are learning over time how to use the modalities of the mTDS as an assistive technology to improve their abilities to interface with the computer. These results are consistent with the results reported in our previous studies carried out with able-bodied participants. In [17], a similar center-out tapping was performed by four participants with average throughput reported to be in the range 1.0-1.3 bits/s, and path efficiency in the range 65-70%. Since the indices of difficulty ( $ID$ ) are the same in this study, the slightly lower throughput values are explained by a higher movement time ( $MT$ ) while the path efficiency is also worse (58-63%). The reason(s) for these differences in performance can be challenging to find because many differences exist between these studies. For instance, in this study, there are a higher number and average age for the participants, all participants have a physical paralysis, they are not as accustomed to computers, and the fact that different possible strategies inherent to some tasks (e.g. games such as peg solitaire and bubble shooter) can affect their performance metrics regardless of the success in performing their targeted computer interaction (e.g. drag-and-drop). In [19], fifteen able-bodied participants were recruited to perform an email sending task. Although the task is similar to the one in this study, there are some important differences that prevents a direct comparison of the results. For instance, in the experiment reported in [19], the task required the participant to open a browser and access the email client prior to composing and sending an email, thus increasing completion time. However, the text entry error rate remains similar with reported values in the range 18-36% which shows that the speech recognition algorithm performs equally regardless of the presence of physical paralysis. This result is expected since the paralysis of our participants did not affect their ability to speak normally. Considering the

results of all these studies, it is rather convincing that the users need more than three sessions to complete their learning phase for this type of human-computer interaction until there is an observable plateau in performance. We are planning to conduct a future study in which the participants will be asked to practice computer access tasks with the mTDS across many days or weeks, preferably in their home or office environments.

Finally, this study does not compare the performance of mTDS to the ATs currently used by the participants because the focus of this work is to show the feasibility of using mTDS in performing computer access tasks by people with tetraplegia. Although this comparison could provide additional insightful information, the inclusion of a participant's AT (e.g. sip-&-puff, head array, chin joystick, switches for minimal range of motion) would have been challenging not only in developing a software interface for all these ATs but also in providing a meaningful analysis across all participants. Additionally, our tasks include interactions made for a proportional control which would be difficult to adapt to switch-like alternative controllers. However, a similar study is being conducted on the use of mTDS to drive a power wheelchair which includes a comparative analysis of performance with the participants' own ATs. This comparison is facilitated because their ATs are already compatible and designed for the purpose of driving a wheelchair, and the analysis is subject dependent. Similarly, at-home case studies are being planned in future work in which the ease of access to real-world and high-value computer tasks will be assessed by people with tetraplegia and compared to their current AT.

## VI. CONCLUSION

In this pilot study, the feasibility of using the multimodal Tongue Drive System (mTDS) as an assistive technology for human-computer interaction was evaluated by fourteen people with tetraplegia on five computer access tasks: navigating a maze, center-out tapping, playing bubble shooting and peg solitaire games, and sending an email. These tasks relied on different combinations of modalities provided by the mTDS to perform complex human-computer interaction tasks such as: drag-and-drop, discrete commands generated by tongue motion, proportional control with head motion, and speech recognition. A variety of performance metrics was selected to evaluate the ability of the participants to complete these tasks, which includes throughput, reaction time, completion time, error rate, path efficiency, typing speed, among others. Overall, the results showed that the performance of the participants was improving over their three sessions of data collection, which can be interpreted as a learning phase. In order to truly assess the performance of the mTDS for computer access tasks, it is imperative that more time for practice and learning be offered to the participants which will be a point of focus in a future study.

## REFERENCES

- [1] "World health organization: Spinal cord injury." [www.who.int/news-room/fact-sheets/detail/spinal-cord-injury](http://www.who.int/news-room/fact-sheets/detail/spinal-cord-injury), 2013. Accessed: 2020-02-17.

- [2] "Christopher reeve foundation: Costs of living with sci." [www.christopherreeve.org/living-with-paralysis/costs-and-insurance/costs-of-living-with-spinal-cord-injury](http://www.christopherreeve.org/living-with-paralysis/costs-and-insurance/costs-of-living-with-spinal-cord-injury). Accessed: 2020-02-17.
- [3] "Policy data brief: Paralysis in the us." [s3.amazonaws.com/reeve-assets-production/RFPParalysisintheUSBrief.pdf](https://s3.amazonaws.com/reeve-assets-production/RFPParalysisintheUSBrief.pdf). Accessed: 2020-02-17.
- [4] T. Santini, W. Fuhl, and E. Kasneci, "Calibme: Fast and unsupervised eye tracker calibration for gaze-based pervasive human-computer interaction," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 2594–2605, Association for Computing Machinery, 2017.
- [5] W. Guo, X. Sheng, H. Liu, and X. Zhu, "Toward an enhanced human-machine interface for upper-limb prosthesis control with combined emg and nirs signals," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 4, pp. 564–575, 2017.
- [6] S. Pancholi and A. M. Joshi, "Electromyography-based hand gesture recognition system for upper limb amputees," *IEEE Sensors Letters*, vol. 3, no. 3, pp. 1–4, 2019.
- [7] A. Skraba, A. Kolozvari, D. Kofjac, R. Stojanovic, E. Semenkin, and V. Stanovov, "Development of cyber-physical speech-controlled wheelchair for disabled persons," in *2019 22nd Euromicro Conference on Digital System Design (DSD)*, pp. 456–463, 2019.
- [8] H. Christensen, I. Casanueva, S. Cunningham, P. Green, and T. Hain, "homeservice: Voice-enabled assistive technology in the home using cloud-based automatic speech recognition," in *Proceedings of the Fourth Workshop on Speech and Language Processing for Assistive Technologies*, pp. 29–34, 2013.
- [9] C. Lin, C. Chiu, A. K. Singh, J. King, L. Ko, Y. Lu, and Y. Wang, "A wireless multifunctional ssvp-based braincomputer interface assistive system," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 11, no. 3, pp. 375–383, 2019.
- [10] F. Arrichiello, P. Di Lillo, D. Di Vito, G. Antonelli, and S. Chiaverini, "Assistive robot operated via p300-based brain computer interface," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6032–6037, 2017.
- [11] B. Jiang, J. Kim, and H. Park, "A new approach of minimizing midas touch problem for a tracer-free tongue-controlled assistive technology," *IEEE Sensors Journal*, 2020. Early Access.
- [12] L. N. S. Andreasen Struijk, B. Bentsen, M. Gaihede, and E. R. Lontis, "Error-free text typing performance of an inductive intra-oral tongue computer interface for severely disabled individuals," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 2094–2104, 2017.
- [13] T. S. Saponas, D. Kelly, B. A. Parviz, and D. S. Tan, "Optically sensing tongue gestures for computer input," in *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, pp. 177–180, ACM, 2009.
- [14] M. Ghovanloo, M. N. Sahadat, Z. Zhang, F. Kong, and N. Sebkhi, "Tapping into tongue motion to substitute or augment upper limbs," in *Micro-and Nanotechnology Sensors, Systems, and Applications IX*, vol. 10194, p. 1019413, International Society for Optics and Photonics, 2017.
- [15] M. N. Sahadat, N. Sebkhi, D. Anderson, and M. Ghovanloo, "Optimization of tongue gesture processing algorithm for standalone multimodal tongue drive system," *IEEE Sensors Journal*, 2018.
- [16] M. N. Sahadat, A. Alreja, N. Mikail, and M. Ghovanloo, "Comparing the use of single versus multiple combined abilities in conducting complex computer tasks hands-free," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 9, pp. 1868–1877, 2018.
- [17] M. Sahadat, N. Sebkhi, and M. Ghovanloo, "Simultaneous multimodal access to wheelchair and computer for people with tetraplegia," in *International Conference on Multimodal Interaction (ICMI)*, pp. 1–7, ACM, 2018.
- [18] N. Sahadat and M. Ghovanloo, "Efficacy assessment of multimodal tongue drive system in comparison to keyboard and mouse," *Archives of Physical Medicine and Rehabilitation*, vol. 98, no. 12, pp. e163–e164, 2017.
- [19] M. N. Sahadat, A. Alreja, and M. Ghovanloo, "Simultaneous multimodal pc access for people with disabilities by integrating head tracking, speech recognition, and tongue motion," *IEEE transactions on biomedical circuits and systems*, vol. 12, no. 1, pp. 192–201, 2018.
- [20] J. Kim, H. Park, J. Bruce, E. Sutton, D. Rowles, D. Pucci, J. Holbrook, J. Minocha, B. Nardone, D. West, *et al.*, "The tongue enables computer and wheelchair control for people with spinal cord injury," *Science translational medicine*, vol. 5, no. 213, p. 213ra166, 2013.
- [21] H. A. Caltenco, B. Breidegard, and L. N. Andreasen Struijk, "On the tip of the tongue: Learning typing and pointing with an intra-oral computer interface," *Disability and Rehabilitation: Assistive Technology*, vol. 9, no. 4, pp. 307–317, 2014.
- [22] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement.," *Journal of experimental psychology*, vol. 47, no. 6, p. 381, 1954.
- [23] R. W. Soukoreff and I. S. MacKenzie, "Towards a standard for pointing device evaluation, perspectives on 27 years of fitts law research in hci," *International journal of human-computer studies*, vol. 61, no. 6, pp. 751–789, 2004.
- [24] R. W. Soukoreff and I. S. MacKenzie, "Measuring errors in text entry tasks: an application of the levenshtein string distance statistic," in *CHI'01 extended abstracts on Human factors in computing systems*, pp. 319–320, ACM, 2001.