A Multimodal Human Computer Interface Combining Head Movement, Speech and Tongue Motion for People with Severe Disabilities

M.N Sahadat, Arish Alreja, Pooja Srikrishnan and Maysam Ghovanloo
School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA
Email:{m.n.sahadat, arish.alreja, psrikrishnan3, mgh}@gatech.edu

***Abstract*—Assistive technologies (ATs) play a crucial role in the lives of individuals with severe disabilities by enabling them to have greater autonomy in performing daily tasks. The Tongue Drive System (TDS) developed at the Georgia Tech Bionics Lab is such an AT, empowering people with severe Spinal Cord Injuries (SCIs) to be more independent. Earlier versions of the TDS have offered tongue motion and speech as means of driving mouse activity and keyboard input. In this paper, we introduce a new multi-modal Tongue Drive System (mTDS), which incorporates head tracking to deliver proportional control of a mouse cursor. The mTDS integrates this new capability while preserving tongue motion and speech from previous versions and offers a richer means of driving computing interfaces, than previously available to individuals with severe disabilities. In experimental trials, 3 able bodied subjects attempted to initiate, dictate and send an email using the mTDS. The mean task completion times of expert, intermediate and novice users with mTDS were 1.48, 4.46 and 6.02 times compared to them using a keyboard and mouse interface. We also observed improved user efficacy with repeated use of the mTDS.**

Keywords: Assistive technologies, multi-modal Tongue Drive System, speech recognition, head tracking, proportional head control

I. Introduction

There are about 276,000 people in the United States living with Spinal Cord Injuries (SCIs) and about 12,500 new cases are reported every year [1]. Assistive Technologies (ATs) are a means of enhancing independent mobility, enabling effective communication and computer access for these individuals. By offering these capabilities, ATs can unlock greater autonomy in accomplishing daily tasks, mitigate health care costs and increase the quality of life for individuals with severe disabilities.

Non-Invasive Brain Computer Interfaces (BCIs) that capture electrical activity with the aim to decode the intended actions of the user are an appealing approach for ATs. Methods like electroencephalogram (EEG) and electrooculogram (EOG) have been explored as ATs [2],[3]. Developing chronically wearable systems with these methods remains a challenge because of user discomfort involved with managing electrode placement, as well as the Midas Touch problem [4], which makes it difficult for the system to differentiate between an intended user action and normal cortical activity.

This work is supported by NSF awards CBET-0828882 and IIS-0803184.

The tongue offers a promising alternative with its sophisticated motor control capabilities and higher degrees of freedom. Indeed, several tongue-computer interfaces such as Tongue-Touch Keypad, Tongue Mouse and Tongue Point have been proposed [5],[6]. The Tongue Drive System (TDS) is a minimally invasive wireless AT. It tracks a user's tongue position using a magnetic tracer that is attached to the tip of the tongue as a piercing. Magnetic sensor arrays mounted on a headset run along either side of the user's cheek as shown in Fig. 1 and measure variations in the magnetic field as the tongue moves. Machine learning algorithms reconcile the sensor data to triangulate tongue position, mapping specific tongue positions to user defined commands [7]. Previous studies [8],[9],[10] have demonstrated the efficacy of the TDS in driving computing interfaces. Subsequent introduction of a dual-Mode Tongue Drive System (dTDS) [11] integrated speech recognition to accept keyboard input in conjunction with tongue motion for mouse control. dTDS enabled higher throughput in tasks involving both text entry and selection, with its combination of modalities optimized for each task category. Studies have also shown that, head orientation for restoring cursor control in individuals with SCI, provides better accuracy, precision and shows less dependence on target direction [12].

Encouraged by the demonstrated performance gains of the dTDS and motivated by the desire to maximize the efficacy of SCI patients for tasks that focus on computer interfaces, we introduce a new multi-modal Tongue Drive System (mTDS) in this paper. The new mTDS integrates tracking head orientation, tongue motion and speech to drive mouse cursor movement, mouse clicks and keyboard input on computing interfaces in an intuitive and seamless manner.

II. Hardware Architecture

The mTDS hardware consists of a magnetic sensor array, an inertial sensor unit, a commercial grade bluetooth microphone, and the main board which contains all the electronic components. The main board is housed in a 3D printed plastic box mounted on a commercial grade headset illustrated in Fig. 1. The hardware build breaks down into the following three high level sub-systems which are discussed in more detail below:

A. Sensing Subsystem

Two printed circuit boards (PCB) $38.35\ mm \times 7.11\ mm$ house the magnetometer array that measure the change of

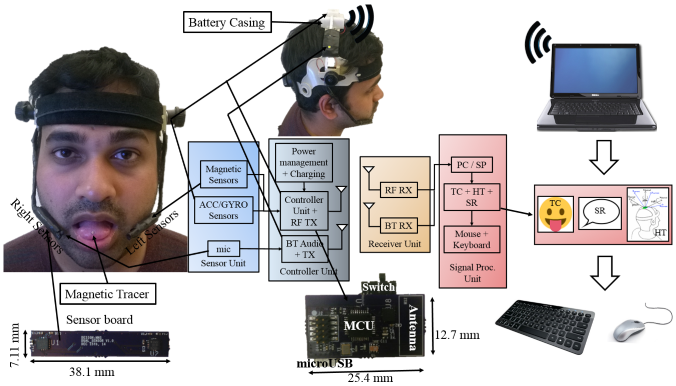

Fig. 1. A physical, block and board level diagram of the multimodal Tongue Drive System (mTDS) headset.

magnetic field resulting from the magnet tracer because of the tongue motion. Each sensor board has two 3 - axis magnetometers and two 3 - axis accelerometer (LSM303D, STMicroelectronics, Switzerland). These sensor boards connect to the main board via a 10 pin, $0.5\ mm$ pitch, flat flexible cable (Samtec, New Albany, IN).

B. Control Unit

The Control Unit houses a power supply unit, a charging unit, serial peripheral interface (SPI) connections from the sensor PCBs and a transceiver MCU. All these components are compactly fit on a $25.4\ mm \times 12.7\ mm$ custom designed PCB which also houses a 2.4 GHz inverted F wiggled PCB antenna for data transmission. The power supply unit contains a LDO (TPS71730) that converts a battery voltage of $4.5\ V$ to a stable low noise $3\ V$ power supply. The power supply powers both the transmitter board and sensor arrays. The on-board battery is charged using a microUSB port and driven by a linear charging IC (LTC4054). The RF transceiver controller (CC2510) receives data from the four 3 - axis magnetometer sensors over an SPI connection at a baud rate of 115.2 Kbps. The magnetometers are configured to operate at a sampling frequency of 100 Hz representing each data sample as an unsigned 16 bit integer. An additional inertial sensor IC (MPU 9250, Invensense, San Jose, CA, USA) is housed on the board and captures 3 axes each of accelerometer, gyroscope and magnetometer data at a sampling frequency of 92 Hz in the form of 16 bit unsigned integers. This data is used to determine head orientation. The same SPI port used to receive magnetometer data receives data from the inertial sensors using multiple chip selects. The data from all sensors is composed together in a 40 byte packet and transmitted to a USB Dongle (connected to the target PC) at a baud rate of 500 Kbps.

C. USB Dongle

The USB Dongle has a RF transceiver (CC2510), a USB controller (CY7C64225) and a chip antenna. The dongle receives data from the mTDS headset and transmits it to the PC via USB at a baud rate of 921.6 Kbps.

III. Software implementation

A graphical user interface (GUI) is implemented in LabView (National Instruments, Austin, TX, USA) to facilate user interaction with the mTDS system. The main GUI takes the user through a workflow which begins with ensuring that the mTDS headset in use can send data to the PC.Then the user goes through a series of calibration routines to calibrate the software from the sensors of the mTDS headset in use. This includes determining sensor orientations, calibrating noise levels, training the tongue tracking machine learning algorithms by completing specific tongue position drills and training the speech recognition engine to recognize the pitch, tone and accent of the user by reading out pre-defined paragraphs of text [11]. After completing these calibration procedures, the integrated mTDS is ready for use. The software creates profiles where it stores important calibration properties which can be retrieved for future use. The software stack is composed of 3 key subsystems which are discussed below:

A. Tongue Commands

The position of the tongue is tracked in real time using data from the magnetic sensor array, using the same SVM classifier described in the previous version of the TDS[7].

B. Speech Recognition (SR)

The mTDS currently integrates the Dragon Naturally Speaking v13 (Nuance, Burlington, MA, USA) speech recognition engine owing to its low cost and multi-platform support. The mTDS software architecture however, retains the flexibility to use other commercially available speech recognition engines.

C. Proportional Head Control

The Proportional Head Control routine gives a user the ability to move a mouse cursor in any direction and at any speed by tilting their head. This capability can deliver the usability of modern computer mice to individuals with disabilities. The routine is written in MATLAB (Mathworks Inc, Natick, MA USA) and integrated into LabView for real-time operation using MathScriptRT blocks. The algorithm relies on the MPU 9250 IC for sensory input. Fig. III-C depicts a functional block diagram of the proportional head control algorithm. The blocks are grouped into three categories.

1) Calibrate Sensor Data: The mTDS tracks Pitch and Roll only. Magnetometer data from the MPU 9250, needed for calculating Yaw is unused. Calibrating 6 axes of accelerometer ($a_{x,y,z}$) and gyroscope ($r_{x,y,z}$) sensor readings from the MPU 9250 involve three transformations.

Firstly, the sensor readings reported as 16 bit unsigned integers [0-65535] at 92 Hz need to be scaled to the dynamic ranges of the MPU 9250 accelerometer($\pm 2\ g's$) and gyroscope($\pm 250°/second$) per (1) and (2).

$$a_{x,y,z} = \frac{Accelerometer\ Range \times (a_{x,y,z} - Offset)}{Scale\ Factor} \quad (1)$$

$$r_{x,y,z} = \frac{Gyroscope\ Range \times (r_{x,y,z} - Offset)}{Scale\ Factor} \quad (2)$$

The $Offset$ value of 32768 represents the zero value for sensor readings. The $Scale\ Factor$ of 32768 relates the maximum magnitudes of sensor readings to the physical values they represent ($AccelerometerRange = 2\ g's,\ GyrosocopeRange = 250°/second$).

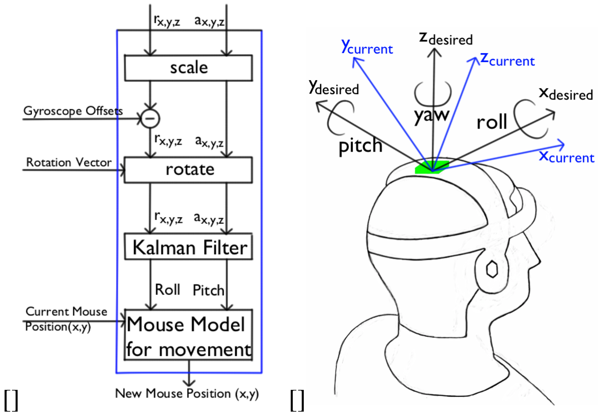

Fig. 2. (a). Functional blocks comprising the Proportional Head Control routine (within blue box) and its I/O interactions with the mTDS LabVIEW application. The blocks illustrate the sequence in which they transform the sensor data into mouse movement. (b). MPU 9250 IC reference axes (blue) may differ from physical axes (black) for the user based on how the MPU (green box) is mounted on the mTDS chassis.

Secondly, the gyroscope offsets are subtracted from the scaled $r_{x,y,z}$ sensor readings from (2). MEMS gyroscopes like ones used by the MPU 9250 rely on the Coriolis effect to detect angular rotation. At rest, their readings include DC and low frequency noise that is difficult to distinguish from actual rotation. Combating this noise is essential to prevent accumulation of significant ”Gyroscope Drift” errors in pitch/roll calculations during continuous operation [13].

A one-time calibration procedure determines the DC noise. This is achieved by averaging ˜1 minute of gyroscope data from the MPU 9250 when the headset is at rest and in an upright position.

$$Gyroscope\ Offset\ r_{x,y,z} = \sum_{t=0}^{t=60sec} \frac{r_{x,y,z}[t]}{92\ Hz \times 60\ sec} \tag{3}$$

During normal operation, each reading is corrected as follows.

$$r_{x,y,z} = r_{x,y,z} - Gyroscope\ Offset\ r_{x,y,z} \tag{4}$$

Finally, the accelerometer and gyroscope sensor readings must be rotated to the upright position aligned with the user's reference axes as depicted in Fig. III-C. A vector which can apply this rotation (7) to sensor readings is calculated by keeping the headset stationary in an upright position with a known/desired acceleration vector A_d (5) and determining the current orientation A_c of the MPU (6) by averaging accelerometer data captured during the MPU 9250 gyroscope offset calibration procedure described in the previous step.

$$A_d = 0\hat{i} + 0\hat{j} - 1\hat{k} \tag{5}$$

$$A_c = A_x\hat{i} + A_y\hat{j} + A_z\hat{k} \tag{6}$$

The rotation vector is then determined as follows.

$$GG = \begin{bmatrix} A_c \cdot A_c & -|A_c \times A_d| & 0 \\ |A_c \times A_d| & A_c \cdot A_d & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$FF = \begin{bmatrix} A_c & \frac{A_d - A_c \cdot A_d \times A_c}{|A_d - A_c \cdot A_d \times A_c|} & A_d \times A_c \end{bmatrix}$$

$$Rotation\ Vector = FF \times GG \times FF^{-1} \tag{7}$$

During normal operation, the rotation vector is applied as follows.

$$a_{x,y,z} = Rotation\ Vector \times a_{x,y,z} \tag{8}$$

$$r_{x,y,z} = Rotation\ Vector \times r_{x,y,z} \tag{9}$$

2) Deduce Head Orientation:
Pitch and Roll control mouse movement. They can be determined either using accelerometer reading only (10), (11) or integrating gyroscope readings over time.

$$Pitch = \tan^{-1}(\frac{-a_x}{a_z}) \times \frac{180^\circ}{\pi} \tag{10}$$

$$Roll = \tan^{-1}(\frac{a_y}{a_x^2 + a_z^2}) \times \frac{180^\circ}{\pi} \tag{11}$$

Both methods have limitations. Accelerometer readings inherently include high frequency noise and are also vulnerable to sudden jerky movements. Gyroscope readings are affected by ”Gyroscope Drift”[13] mentioned earlier. Kalman filters are an attractive method of fusing 2 signals with different levels of noise together to achieve a resultant signal which has lesser noise than either of the component signals [14]. Commonly used in inertial navigation, they are an appropriate choice here to fuse accelerometer and gyroscope sensor readings and determine robust Pitch and Roll signals for tracking head orientation. Equations (12) and (13) describe the state model.

$$x_t = A \times x_{t-1} + w_{t-1} \tag{12}$$

$$z_t = H \times x_t + v_t \tag{13}$$

Where $A = \begin{bmatrix} 1 & \Delta\ t \\ 0 & 1 \end{bmatrix}$, $x = \begin{bmatrix} Pitch & Roll \\ r_y & r_x \end{bmatrix}$,H is a 2×2 Identity matrix, Δt is the sampling period and w and v represent the process and measurement noise, respectively.

3) Translate Head Orientation to Mouse Movement:
The mouse movement model converts angular displacement of the head along the Roll and Pitch axes to mouse cursor velocities along the X and Y axes. Residual head movement when the user intends to stay in resting position can result in small undesired movement of the mouse cursor. To provide robustness against residual movement, a threshold function ($\lambda_{roll,pitch}$) is incorporated. Head movement capabilities of individual users can vary significantly based on the severity of their injuries. A Gain parameter α enables customization of movement sensitivity (slope of transfer function) based on an individual's movement capabilities. The transfer function is described by (14) and visualized in Fig. 3.

$$\begin{bmatrix} Mouse_{x|t} \\ Mouse_{y|t} \end{bmatrix} = \begin{bmatrix} Mouse_{x|t-1} \\ Mouse_{y|t-1} \end{bmatrix} + \alpha \begin{bmatrix} Roll_t - \lambda_{roll} \\ Pitch_t - \lambda_{pitch} \end{bmatrix} \tag{14}$$

IV. Experimental Study

Three subjects (age: 23-32, 2 male and 1 female) who were all non-native English speakers, participated in this study. The task required composing and sending an email with specific content (60 words, 379 letters) to a given email address. The contents and target email address remained fixed across trials. Each subject conducted 5 trials with the mTDS and the keyboard/mouse interfaces. The task completion times were recorded for all trials and are analyzed in the Results section.

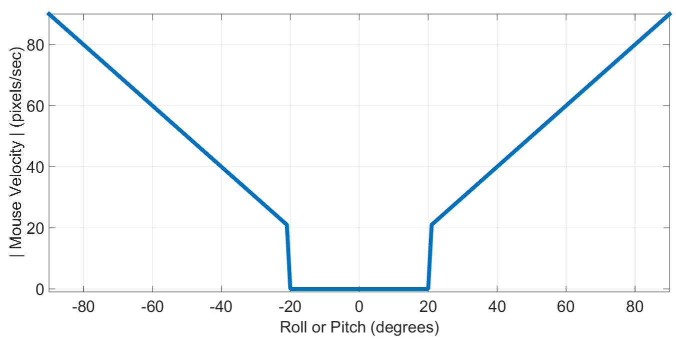

Fig. 3. Transfer function translating angular displacement to mouse cursor velocity. Threshold $\lambda_{roll,pitch} = 20°$ and $\alpha = 1$. Roll and Pitch are symmetrical in this case and have the same transfer function.

TABLE I. TASK COMPLETION TIME (MEAN $\pm$ STD): mTDS vs. KEYBOARD/MOUSE (KnM)

EXP.	SUBJECT A	SUBJECT B	SUBJECT C
mTDS duration (min)	3.98 ± 0.42	7.41 ± 2.96	7.23 ± 4.37
KnM duration (min)	2.69 ± 0.37	1.23 ± 0.08	1.62 ± 0.31

V. Results

Each of the three subjects in the study had different levels of prior exposure to the mTDS. Subject A was an experienced user with prior exposure to the tongue commands and the Dragon Natural Speaking platform. Subject B was a novice with no exposure to mTDS. Subject C was only familiar with the speech recognition capabilities of the mTDS. Each subject shows a trend of improved performance with repeated use of the mTDS shown in Fig. 4. The trial over trial performance improvements achieved by all subjects using mTDS are different from the performance trends when using keyboard/mouse, where the performance is relatively invariant compared to the mTDS trials. Because all subjects use Keyboard and Mouse interfaces on a daily basis. The comparison between the typing duration using the mTDS and keyboard and mouse are shown in Table I.

VI. Conclusion

We have demonstrated that a multi-modal Tongue Drive System integrating head movement, speech and tongue motion can be used to accomplish complex tasks such as email communication over a standard PC. Our preliminary trials reveal that the user efficacy improves with repeated use. The experiment required repeating the same email 5 times, which as a limitation may influence typing speed as a result of short term memorization of the email content for subjects. Our future plans include a more extensive performance evaluation of the mTDS interface with standardized HCI tasks like center out and more complex integrative computing tasks like word processing and spreadsheet applications. Future experimental trials will also include a more significant subject population comprised of individuals with and without disabilities. Additionally, we plan to explore refinements which can improve performance. These include adapting the standard mouse ballistics transfer functions for the mTDS; solving the 'Midas touch'[4] problem to ensure end-users can seamlessly switch between exercising tongue commands and speech recognition without interference among the modalities; and more advanced

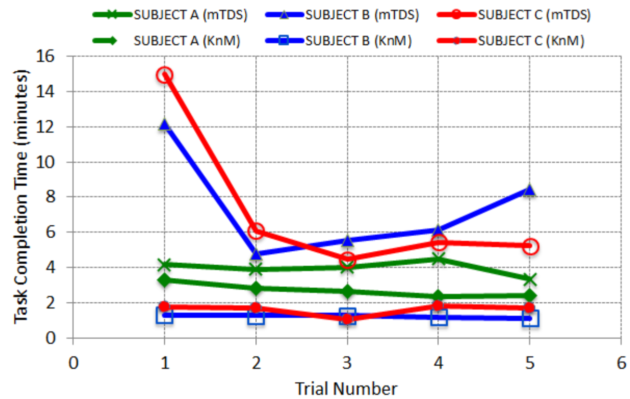

Fig. 4. The performance evaluation over time: mTDS vs. Keyboard/Mouse (KnM).

control primitives in the future such as cut, copy, paste, right-click, alt-tab to switch between applications.

References

[1] National Spinal Cord Injury Statistical Center, "Spinal Cord Injuries Facts & Figures at Glance," 2013, [Online], Available: http://www.msktc.org/lib/docs/Data˙Sheets˙/SCIMS˙FactsFigures˙Feb2015.

[2] Barea R, et al., "System for assisted mobility using eye movements based on electrooculography" in IEEE Trans Neural Syst Rehabil Eng. 2002; 10(4):209-18.[PMID:12611358].

[3] Wolpaw JR, et al., "Brain-computer interfaces for communication and control" Clin Neurophysiol. 2002; 113(6):767-91.[PMID:12048038].

[4] M. M. Moore, "Real-world applications for brain-computer interface technology" in IEEE Trans. Rehabil. Eng, vol. 11, no. 2, pp. 162-165, 2003.

[5] Lau C, O'Leary S., "Comparison of Computer Interface devices for persons with severe physical disabilities" Am J Occup Ther. 1993; 47(11):1022-30. [PMID:8279497].

[6] Salem C, Zhai S, "An isometric tongue pointing device" in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 1997 Mar 22-27; Atlanta, Georgia. New York(NY): Association for Computing Machinery: 1997.p. 22-27.

[7] X. Huo, et al., "A magneto-inductive sensor based wireless tongue-computer interface," IEEE Trans. on Neural Sys. Rehab. Eng., vol. 16, no. 5, pp. 497-504, Oct. 2008.

[8] Huo X, Ghovanloo M., "Evaluation of a wireless wearable tonguecomputer interface by individuals with high-level spinal cord injuries" in J Neural Engineering. 2010 Apr;7:026008.

[9] Yousefi B, et al., "Using Fittss law for evaluating tongue drive system as a pointing device for computer access" in IEEE EMBC 2010 Sep;:44044406.

[10] B. Yousefi , et al., "Quantitative and comparative assessment of learning in a tongue-operated computer input device" in IEEE Trans. Info. Tech. Biomed., vol. 15, no. 5, pp. 747-757, 2011.

[11] Xueliang Huo, et al., A Dual-Mode Human Computer Interface combining Speech and Tongue motion for people with severe disabilities in IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 21, no. 6.pp.979-990.

[12] M. R. Williams and R. F. Kirsch, "Evaluation of head orientation and neck muscle EMG signals as command inputs to a human-computer interface for individuals with high tetraplegia" in IEEETrans. Neural Syst. Rehabil. Eng., vol. 16, no. 5, pp. 485-496, 2008.

[13] Abyarjoo, et al. "Implementing a sensor fusion algorithm for 3D orientation detection with inertial/magnetic sensors." Innovations and Advances in Computing, Informatics, Systems Sciences, Networking and Engineering. Springer International Publishing, 2015. 305-310.

[14] Faragher, R., "Understanding the Basis of the Kalman Filter Via a Simple and Intuitive Derivation [Lecture Notes]," Signal Processing Magazine, IEEE , vol.29, no.5, pp.128,132, Sept. 2012.