Published in final edited form as:
Conf Proc IEEE Eng Med Biol Soc. 2011 ; 2011: 6393–6396. doi:10.1109/IEMBS.2011.6091578.

Using Speech Recognition to Enhance the Tongue Drive System Functionality in Computer Access

Xueliang Huo [Student Member, IEEE] and **Maysam Ghovanloo* [Senior Member, IEEE]**
GT-Bionics Lab, School of Electrical and Computer Engineering at Georgia Institute of Technology, Atlanta, GA, USA

Abstract

Tongue Drive System (TDS) is a wireless tongue operated assistive technology (AT), which can enable people with severe physical disabilities to access computers and drive powered wheelchairs using their volitional tongue movements. TDS offers six discrete commands, simultaneously available to the users, for pointing and typing as a substitute for mouse and keyboard in computer access, respectively. To enhance the TDS performance in typing, we have added a microphone, an audio codec, and a wireless audio link to its readily available 3-axial magnetic sensor array, and combined it with a commercially available speech recognition software, the Dragon Naturally Speaking, which is regarded as one of the most efficient ways for text entry. Our preliminary evaluations indicate that the combined TDS and speech recognition technologies can provide end users with significantly higher performance than using each technology alone, particularly in completing tasks that require both pointing and text entry, such as web surfing.

I. Introduction

Tongue Drive System (TDS) is a minimally invasive, unobtrusive, wireless, wearable tongue operated assistive technology (AT), which can enable individuals with severe physical disabilities to access computers and drive powered wheelchairs through their volitional tongue movements [1]-[3]. TDS detects a set of user-defined tongue gestures or positions inside the mouth by tracking a magnetic tracer secured on user's tongue with an array of magnetic sensors mounted on a wearable headset. It then associates these movements with specific control commands that can be used to emulate the mouse and joystick functions for computer access and wheelchair navigation, respectively [1], [2].

The performance of the TDS as a pointing device for computer access has been quantitatively and comparatively evaluated (vs. keypad) and reported in [4] and [5]. In our first clinical trial we also demonstrated that the TDS potential end users, i.e. individuals with high level spinal cord injuries, can use this system to substitute the mouse function in moving the mouse cursor to complete computer access tasks or replace a manual joystick to drive a powered wheelchair [3]. These experiments have proven that the TDS can independently provide its users with an effective mean to control their environments and access computers.

It is well understood that any AT that is designed around only one method of input may not be fast and flexible enough to meet the diverse needs of the end users in today's hectic and demanding life conditions [6]. A multimodal device that expands the physical access beyond one input channel, on the other hand, can potentially improve the speed of input by increasing the information transfer bandwidth between users and computers [7], [8]. In

mgh@gatech.edu.

addition, multimodal interfaces increase the number of alternative available to users to accomplish a task, thus give users the ability to switch among different input modalities, based on their convenience, familiarity, and environmental conditions [9]. Multimodal devices can give users a choice of how to do their tasks. These devices also provide their users with more options to cope with the fatigue, which is an important factor that affects the acceptability of ATs, and therefore can result in greater user satisfaction and technology adoption.

The TDS, in the current form, which offers its users with six simultaneously accessible commands, is mainly designed to substitute mouse cursor movement and click functions in computer access. Even though TDS can provide full typing feature when combined with an on-screen keyboard, its fast response time (< 0.5 s [1]) and the relatively limited number of discrete commands (compared to a keyboard) makes it more suitable for mouse cursor control as opposed to typing. On the other hand, speech recognition has almost unlimited number of available commands, and regarded as one of the most efficient ways for text entry, which after training can outperform rapid typing with a keyboard in a quiet environment. Individuals with severe disabilities can benefit from this technology as long as their vocal abilities are intact. The speech recognition software also allows its user to navigate the mouse cursor using a set of predefined voice commands with relatively long response time because they need short pauses before and after issuing each command.

Therefore, combining TDS and speech recognition can potentially offer the users the best of both in following ways: 1) Increase in speed since each device can be used for its target functions; 2) Allowing users to select either technology to use depending on the operating and environmental conditions, such as fatigue and noise, respectively [7].

The main objective of the presented work is to enhance the functionality of the TDS in computer access by adding a speech recognition input channel to the current system, and turn it into a multimodal and multi-function AT that can be used for a wide variety of tasks related to computer access and environmental control. In this system, both TDS and speech recognition technologies are simultaneously accessible to the users, particularly for mouse navigation and typing, respectively. Users have the flexibility to choose the device they want to use for any specific task without external assistance. For example, in a quiet indoor environment, using TDS for moving the mouse cursor and the speech recognition for text entry works well. However, in a noisy environment, the user might prefer to use TDS for both tasks.

II. Multimodal Tongue Drive System

The multimodal Tongue Drive System (mTDS) is an enhanced version of the TDS with an add-on one-way wireless audio link to acquire and transmit users' vocal commands. mTDS consists of four major components: 1) a small permanent magnetic tracer attached on the tongue by tissue adhesive or piercing; 2) a wireless headset containing an array of three-axial magnetic sensors to detect the magnetic field generated by the magnetic tracer, a miniaturized microphone incorporated with the left magnetic sensor module to capture the user's voice, and a wireless control unit to record and transmit the sensor and audio samples; 3) a USB receiver operating at same RF frequency as the headset, which wirelessly receives the sensor and audio samples and sends them to a PC or iPhone; and 4) a graphical user interface (GUI) running on the PC or iPhone with embedded sensor signal processing (SSP) algorithm, which recognizes the position of the magnet, hence, the position of the tongue within the oral space. Fig. 1 shows various components of the first mTDS prototype.

A. Permanent Magnetic Tracer

Benefiting from the new highly sensitive and small 3-axial magnetic sensors and a smart SSP, we were able to use disc-shaped NdFeB rare earth magnets (K&J Magnetics, Jamison, PA) with small size (Ø 3 mm × 1.6 mm) and high residual magnetic strength (B_r = 14,500 Gauss) as the tracer. Using small tracers is desired to reduce possible discomfort resulted from the magnet attachment, while the higher B_r can compensate for the signal-to-noise (SNR) degradation in the magnetic sensor output due to shrinking the magnet size.

B. Wireless Headset

Wireless headset, which block diagram is shown in Fig. 2, is a key component of the mTDS. The headset was equipped with a pair of goosenecks, each of which bilaterally holds two 3-axial anisotropic magneto-resistive (AMR) sensors HMC1043 (Honeywell, Morristown, NJ) near the subjects' cheeks, symmetrical to the sagittal plane.

The sensing element of the AMR sensor is made of nickel-iron thin film, which resistance changes in presence of a magnetic field. This change can be measured using a Wheatstone bridge configuration to characterize both magnitude and direction of the field [10]. In the HMC1043, three AMR orthogonal sensors in X, Y and Z axes measure the magnetic field vector in 3-D. In the mTDS, the differential output signals from each HMC1043 sensor bridge are multiplexed locally on the sensor module, and the outputs from the two modules are further multiplexed on the control unit to yield only one differential voltage pair. These time multiplexed signals are amplified by a low power low noise instrumentation amplifier, INA331 (TI, Dallas, TX), with a gain of 200 V/V. A low-power microcontroller (MCU) with built-in analog-to-digital converter (ADC) and 2.4 GHz RF transceiver (CC2510, TI, Dallas, TX) samples each sensor output at 50 Hz, while turning on only one sensor at a time to save power. Each sensor is duty cycled at 2%, which results in a total duty cycle of 8%. To avoid sensor sensitivity and linearity degradation in the present of strong fields (> 20 Gauss) when the magnetic tracer is very close to the sensor (<1 cm), the MCU generates a 2 μs short pulse to reset the sensor right before the sensor output is sampled. After all four sensor output are sampled, the results are packed into one data frame ready for RF transmission.

The audio signal acquisition was independent of magnetic sensor sampling and performed by an audio codec TLV320- AIC3204 (TI, Dallas, TX), through the built-in inter-IC sound (I2S) interface of the CC2510 MCU. A miniaturized SiSonic MEMS microphone (Knowles, Itasca, IL) was placed near the tip of the left sensor board, as shown in Fig. 1, to continuously capture the sound signal when the user is speaking. The microphone is directly connected to the audio codec on the control unit which has dedicated power supply, ground, and signal wires to minimize the interference from digital control lines. The audio codec is programmed to operate at the lowest performance level with single-ended mono input, 8 kbps sampling rate, and 16 bits of resolution to minimize power consumption. This configuration provided sufficient quality to capture the voice signal in the frequency range of 100 ~ 2000 Hz using the SiSonic microphone with 59 dB SNR.

Digitized audio samples are read by the MCU through I2S and compressed to an 8 bit format using the CC2510 built-in μ-Law compression hardware to save the RF bandwidth. Due to the time critical nature of streaming audio, the audio data transfers within the MCU, from I2S to RAM and from RAM to the RF transmitter, are accomplished using direct memory access (DMA) to minimize the CPU intervention and the resulting latency. Once a completed audio frame (54 samples) has been acquired, in 6.75 ms, the MCU assembles an RF packet containing one audio and one data frame and transmits it wirelessly. Since the audio and data frames are generated at different intervals (6.75 ms vs. 20 ms), only one out

of every three RF packets contains both audio and data samples, and the other two includes only audio samples. These two types of packets are tagged with different preambles so that they can be recognized and properly disassembled on the receiver side.

The power management circuitry includes a pair of AAA Ni-Mn batteries, a voltage regulator, a low voltage detector, and a battery charger. The system consumes ~30 mA at 2.5 V supply and can run for more than 25 hours following a full charge. Table I summarizes some of the key specifications of the first mTDS prototype.

C. USB Receiver

The same type of MCU and audio codec are used on the mTDS wireless receiver, which is responsible for receiving the RF packets and delivering them to the computer. After being extracted from the RF packet, the data samples are directly sent to the computer via USB, while the audio samples are streamed to a playback audio codec through the I2S interface and converted to analog sound signals, which are then applied to the microphone input of the computer.

D. Graphical User Interface (GUI)

Even though the mTDS GUI runs in the LabVIEW environment, its SSP engine has been implemented in C to improve the computational efficiency. The SSP algorithm uses the K Nearest Neighbor (KNN) classifier to identify the incoming sensor samples based on their features, which are extracted through Principal Components Analysis (PCA) from the data that is collected during a training step prior to testing [1]. The current TDS prototype supports six individual commands that are simultaneously available to the user including four directional (LEFT, RIGHT, UP, and DOWN) and two selection commands (LEFT- and RIGHT-CLICK).

Any piece of commercially available or customized speech recognition software that works with a regular microphone can be used with the multimodal TDS, because the audio signals are directly applied to the microphone input of the computer. We have chosen the Dragon Naturally Speaking v10.0 (Nuance, Burlington, MA) since it has been widely used by the disabled community.

III. Preliminary Evaluation

A 30-year old male subject, with Asian ethnicity, who is a member of the research team, performed a web browsing experiment to evaluate the performance of the mTDS in completing realistic computer access tasks that involved both mouse navigation and typing. The subject had prior experience with both TDS and the Dragon, however, he was not a regular user of either technology. The subject was asked to wear the mTDS headset and sit ~1 m from a 22″ monitor with 1280×800 resolution. The subject trained the Dragon Naturally Speaking software by reading 10 short paragraphs provided by the manufacturer. Then he conducted the TDS calibration, tracer attachment, command identification, and training steps as explained in [1] to define his six mTDS tongue commands.

The mouse cursor was initially positioned in the middle of the monitor screen and the subject was required to navigate the cursor to complete the following tasks in the same order, while the computer kept track of the elapsed time and user commands: 1) Open a web browser [Internet Explorer] by clicking its icon in the Windows-XP start menu; 2) Type www.amazon.com in the browsers address bar and click on the [Go] button to reach the Amazon website; 3) Type *wireless mouse* in the search box and click on [Search] button to find the related products; 4) Click on the name of the first item in the list of search results and then click on [Add to Cart] button to add the item to the shopping cart; 5) Click on

[Proceed to checkout] button; 6) Close the browser by clicking on the red cross on the top right side of the browser window. All in all, the subject had to complete a minimum of 15 mouse cursor movements (excluding those for typing with TDS), 9 mouse clicks, and 28 typed-in characters. The subject's activities on the computer screen was recorded using Camtasia Studio (TechSmith, Okemos, MI) and analyzed offline to derive the performance merits, such as typing time, cursor navigation time and total completion time.

The subject was required to complete the task using the TDS without Dragon, Dragon alone, and the mTDS with Dragon. The task was repeated four times for each variation, one for practice followed by three testing trials. When using the TDS, the microphone was turned off to deactivate the Dragon. In this case, the directional TDS commands were used to move the cursor on the screen in four directions and the selection commands were used to issue mouse left-click and double-click. Typing in this case was accomplished by navigating the cursor and clicking on an on-screen keyboard (Click-N-Type, Lake Software). When using the Dragon, the TDS function was disabled by shutting down the LabVIEW GUI. A set of predefined verbal commands, such as *move mouse Left/Right/Up/Down, move mouse slow, much faster, and mouse left/right click*, were used to move the cursor and issue mouse clicks through the dictation. In the multimodal mode, both the TDS and Dragon were active, and the subject was requested to use the TDS for mouse navigation and clicks, and the Dragon for typing.

Fig. 3 depicts the results from testing trials, divided into the typing time, cursor navigation time, and total time for completing the task using three different solutions. We also asked the subject to perform the same task with standard mouse and keyboard to have a reference point. Overall, using the mTDS resulted in the best performance in all aspects. TDS outperformed Dragon in term of cursor navigation time (76 s vs. 234 s), while the Dragon was much faster in typing (18 s vs. 114 s). The subject obviously benefited from using both devices, evident from his minimum total completion time when using the mTDS, which was about 42% and 34% of that of using TDS alone and Dragon alone, respectively. Interestingly, the cursor navigation time of TDS did not vary much whether it was used alone or with Dragon. Similarly, the typing time with Dragon was basically the same with and without TDS. These results show that TDS and Dragon can be used together and independently without degrading the user's performance with each individual device.

IV. Conclusion

We have developed a multimodal Tongue Drive System (mTDS) with speech recognition capability by adding a small microphone, a low power audio codec, and a wireless audio link to the original TDS to enhance its functionality in computer access. mTDS allows users to operate the mouse cursor using their tongue motion and type or edit text using speech. Preliminary results supported the idea that a multi- modal AT can significantly improve the speed of completing complex computer access tasks, such as web surfing, where both text entry and cursor navigation are necessary. It was also demonstrated that using TDS with speech recognition does not affect the user's performance with either one of these technologies. We are working to add more input channels, such as head control, to the current mTDS platform to further improve its speed, usability, and end user coverage. We also intend to evaluate the mTDS performance by those with severe disabilities in home/office/outdoors environments.

Acknowledgments

This work was supported in part by the National Institute of Biomedical Imaging and Bioengineering grant 1RC1EB010915 and the National Science Foundation award CBET-0828882 and IIS-0803184.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

References

1. Huo X, Wang J, Ghovanloo M. A magneto-inductive sensor based wireless tongue-computer interface. IEEE Trans Neural Syst Rehabil Eng. Oct.2008 16:497–504. [PubMed: 18990653]
2. Huo X, Ghovanloo M. Using unconstrained tongue motion as an alternative control surface for wheeled mobility. IEEE Trans on Biomed Eng. Jun.2009 56:1719–1726.
3. Huo X, Ghovanloo M. Evaluation of a wireless wearable tongue–computer interface by individuals with high-level spinal cord injuries. J Neural Engineering. Apr.2010 7:026008.
4. Yousefi B, Huo X, Ghovanloo M. Using Fitts's law for evaluating tongue drive system as a pointing device for computer access. IEEE Eng in Med and Biol Conf. Sep.2010 :4404–4406.
5. Yousefi B, Huo X, Veledar E, Ghovanloo M. Quantitative and comparative assessment of learning in a tongue-operated computer input device. Submitted to IEEE Trans Info Tech in Bio Med.
6. Keates S, Robinson P. The use of gestures in multimodal input. Proc of the 3rd Intl ACM Conf on Assist Tech. Apr.1998 :35–42.
7. Smith A, Dunaway J, Demasco P, Peichl D. Multimodal input for computer access and alternative communication. Proc of the 2nd Annual ACM Conf on Assist Tech. Apr.1996 :80–85.
8. Shein, F.; Brownlow, N.; Treviranus, J.; Pames, P. Proc of the Visions Conf: Augmentative and Alternative Comm in the Next Decade. University of Delaware/Alfred I. duPont Institute; Wilmington, DE: 1990. Climbing out of the rut: The future of inter- face technology.
9. Baljko M. The contrastive evaluation of unimodal and multimodal interfaces for voice output communication aids. Proc of the 7th Intl Conf on Multimodal Interfaces. Oct.2005 :301–308.
10. Caruso MJ, Bratland T, Schneider RW, Smith CH. A new perspective on magnetic field sensing. Proc of Sensors Expo. Oct.1998 :195–213.

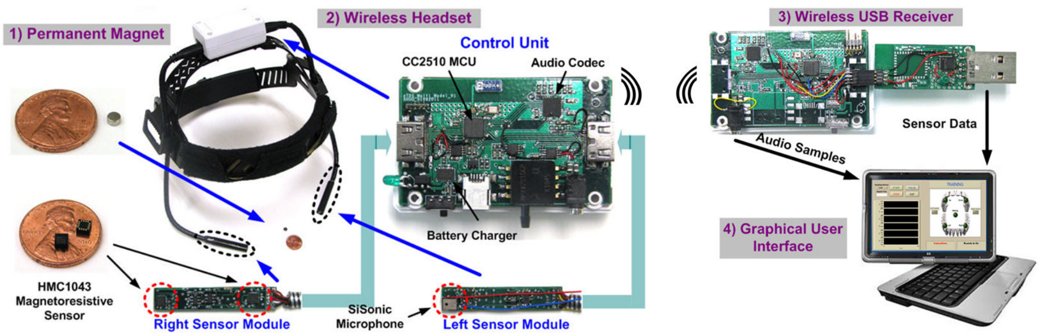

Fig. 1.
Various components of the multimodal Tongue Drive System (mTDS) prototype.

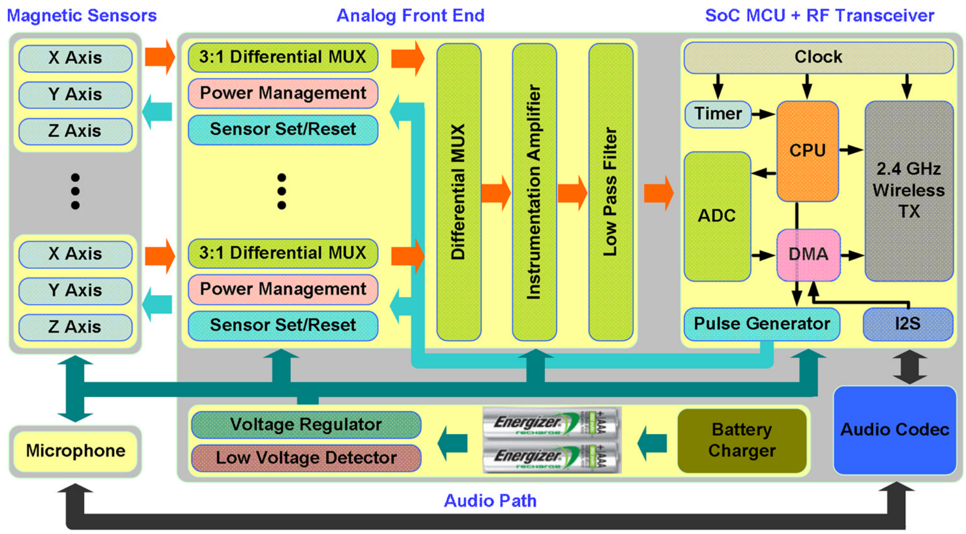

Fig. 2.
Block diagram of multimodal Tongue Drive System headset

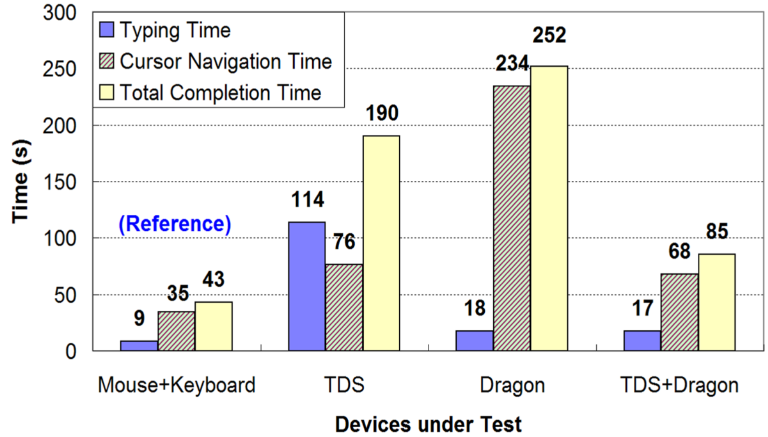

Fig. 3.
Evaluation results of using the TDS, Dragon, and multimodal TDS (mTDS) to complete a web browsing task.

NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

TABLE I

Multimodal Tongue Drive System Hardware Specifications

Specification	Value
Magnetic Tracer	
Material	NdFeB rare-earth magnet
Size (diameter and thickness)	Ø 3 mm × 1.6 mm
Residual magnetic strength	14500 Gauss
Magnetic Sensors	
Type	Honeywell HMC1043 AMR sensor
Dimensions	$3.0 \times 3.0 \times 1.5$ mm^3
Sensitivity/range	1 mV/V/Gauss/± 600 μT
Microphone	
Type	SiSonic SPM0408HE5H
Dimensions	$4.7 \times 3.8 \times 1.1$ mm^3
Sensitivity/SNR	−22 dB/59 dB
Control Unit	
Microcontroller	Chipcon (TI) – CC2510 with built-in RF transceiver
Wireless frequency/data rate	2.42 GHz/500 kbps
Sampling rate	50 sample/s/sensor
Number of sensors/duty cycle	4/8%
Audio codec/interface	TLV320AIC3204/I2S
Audio sampling rate/resolution/compression	8 ksps/16 bits/μ-Law
Operating voltage/current	2.5 V/~ 30 mA (audio on) ~ 6.5 mA (audio off)
Dimensions	$73 \times 40 \times 26$ mm^3
Weight	14 g without batteries