Learning to Type with the Tip of the Tongue: A Performance Study for a Tongue-Computer Interface

Hector A. Caltenco [1], Shellie Boudreau[1], Romulus Lontis[1], Bo Bentsen[2] and Lotte N.S. Andreasen-Struijk[1]

[1]Center of Sensory-Motor Interaction, Aalborg University
Frederik Bajers Vej 7, 9220 Aalborg, Denmark
Email: hcaltenco@ieee.org
Telephone: (+45) 9635 7575
[2]Dental Clinic Andersen & Bentsen
Vendsysselgade 28, 9000 Aalborg, Denmark

Abstract- **This study is motivated by the need to know the characteristics of the learning processes in tongue-computer interaction and to obtain a useful insight to a better design of the tongue-computer interface for computer text input. Tongue-typing can be a good alternative to hand input methods for physically disabled individuals or tasks where hand-typing is not possible. In order to evaluate the process of typing with the tip of the tongue, eight volunteers participated in tip-of-tongue selectivity training experiments using an inductive tongue-computer interface. Performance data based on typing speed and accuracy fits a general learning model based on the power law of practice, which can be used to estimate further improvements of tongue-typing performance. Simulated expert typing rates predict a tongue-typing performance 8 times slower than normal QWERTY keyboard, but duplicate the performance of other alternative input interfaces. Our results encourage the use of a tongue-computer interface over other methods for physically disabled individuals.**

I. INTRODUCTION

Text input interfaces are an essential part of current computer systems. The QWERTY keyboard and the standard mouse (with augmentative communication software) dominate desktop computing, but these input interfaces are designed for hand use, which may not always be convenient. There is, however, a population of users with various physical disabilities who are unable to use hand operated input interfaces. These interfaces are also problematic for restrictive environments or tasks that require complete dedication of both hands, for example, driving, piloting or operation of certain machinery.

Most alternative input interfaces are very noticeable or have little transportability and may present other disadvantages depending on the input source, such as neck pain in head control methods, headaches in eye control methods, and interfering signals in speech control methods. Additionally, brain control methods need further research to reach an acceptable level of reliability.

In a study comparing three input interfaces [1], a Tongue-Touch-Keypad (TTK®) from *New Abilities* [2] was preferred by users due to its discretion, even though it was not the most efficient method. The TTK does not exploit the fine motor control of the tongue and the use of pressure sensors located on the palatal plate may fatigue the user and reduce the speed of sensor activation. Other tongue control systems [3, 4] present similar problems.

A new inductive tongue-computer interface (ITCI) [5] developed at Aalborg University, is partly implantable and can incorporate a larger number of sensors. The sensors can be activated by appropriate positioning of the tongue, which reduces the fatigue and increases the speed of sensor selection. Another new interface that detects intra-oral tongue movements is a magnetic tongue-computer interface (MTCI) "Tongue-Drive" [6]. These inductive and magnetic tongue-computer interfaces are promising text input interfaces without the need of applying pressure on the palatal area.

This study is motivated by the need of knowledge about the characteristics of the learning processes in tongue-computer interaction to create a basis for a better design of a tongue-computer interface for computer text input. For these purposes, tip-of-tongue selectivity training experiments were performed in order to evaluate the learning model for tongue-typing performance using the ITCI. These results are compared to other results found for different alternative text input interfaces reported on literature.

II. PERFORMANCE MEASURES IN TEXT INPUT INTERFACES

Effective assessment of text input interfaces requires two levels of evaluation: the *user* and the *system*. The user must control the signal features, and the system must recognize that control and translate it into device control effectively and consistently. User performance is a correlation between the intended command selection and the actual command selection, while system performance is a measure of the amount of information that the system can transmit. [7]

User and system performance may depend on the same factors, for example the number, position and dimension of the targets, in our case the inductive sensors. A greater number of targets could increase *system* performance, since more options provide more information, but can also decrease the *user* performance by decreasing the accuracy. Other factors that affect both types of performances are reaction time, movement time, distance between targets, etc. For example,

978-1-4244-2156-5/08/$25 (c) 2008 IEEE

faster movement can increase *system* performance by permitting a greater number of selections in a trial, but can also decrease accuracy, and therefore *user* performance.

User and system performance can be combined in the same index of performance in order to optimize the overall performance of the product. This can be done as the evaluation of two aspects: 1) *experimental* performance, which is the overall performance of a specific task assessed as speed and accuracy, and 2) *theoretical* performance, which is the amount of information that can be communicated per unit time, assessed as information transfer rate (bits/sec).

This index of performance, also known as throughput (*TP*), has been a fundamental metric in quantifying input interfaces performance. The definition of *TP* is varied in the literature, but due to the International Standard ISO 9241-9, it has converged onto the ratio of index of difficulty (*ID*) to trial completion time (*MT*).

$$TP = \frac{ID}{MT} \tag{1}$$

A. Theoretical performance

Fitts' law [8] has been used as a framework in a big part of text and pointing performance research [9, 10]. It determines (in bits) the information capacity of the human motor system in controlling amplitude of movement (*A*) with specific sensor width (*W*):

$$ID = \log_2\left(\frac{A}{W} + 1\right) \tag{2}$$

B. Experimental performance

The speed for typing tasks is usually reported as words per minute (wpm) or characters per second (cps). These measures are highly interchangeable counting that in English language an average word has 5 characters counting space. Researchers have used these measures to report typing rates of different text input interfaces [1, 11, 12]. Accuracy is more problematic, because we can deal with different types of errors. The basic types of errors include entering an incorrect character (substitution), omitting a character (omission), adding an extra character (insertion), or swapping neighboring characters (transposition). One common way to deal with errors in typing tasks is to report correct words per minute (cwpm), either by forcing the subject to correct their errors during the trial, or by applying error adjustment methods after the tests [9].

III. MATERIALS AND METHODS

Eight able bodied volunteers (5 male and 3 female, mean age of 24.7, SD=3.5) participated in 3 consecutive 30 minutes/day tongue selectivity training sessions. Each session consisted on 44 trials that were performed for intervals of 30 seconds. Three of the subjects also performed an additional tongue selectivity session one week post training.

A. Experimental Setup

Using dental acrylic, nine air cored inductors were placed on a palatal plate resembling the ones used as dental retainers. The palatal plate with the inductors was placed at the hard palate and an activation unit consisting of a small ferromagnetic stainless steel cylinder was used to activate the coil sensors. This activation unit was glued to the tip of the tongue using tissue glue. The subjects activated one of the nine inductive sensors by positioning the tongue in a manner that placed the activation unit in the centre of the different inductor coils. There were two different setups used:

1. For the first setup, 2-3 coils were connected in series, and the detection was performed by using thresholds within the shared channels. This process was used for the first four subjects. The specifications and detailed signal processing of the ITCI using this setup can be found in [13].
2. For the second setup, all the coils were connected in a separate channel and there was no need for thresholding signals to differentiate which sensor was activated. This setup was used for the last four subjects.

Each of the following characters: "ABCDEFGHI" was related to the activation of a sensor and displayed on a computer screen, located in front of the user, when the corresponding inductor was activated (see Figure 1). Each trial consisted on typing one of the 15 different sequences of characters, two of which were test sequences. Test sequences were typed 3 consecutive times at the beginning, middle and end of each 30 min. session (a total of 9 times). The rest of the sequences were typed only once (2 consecutive times) during each 30 min. session. The types of sequences for each training day are listed in Table I.

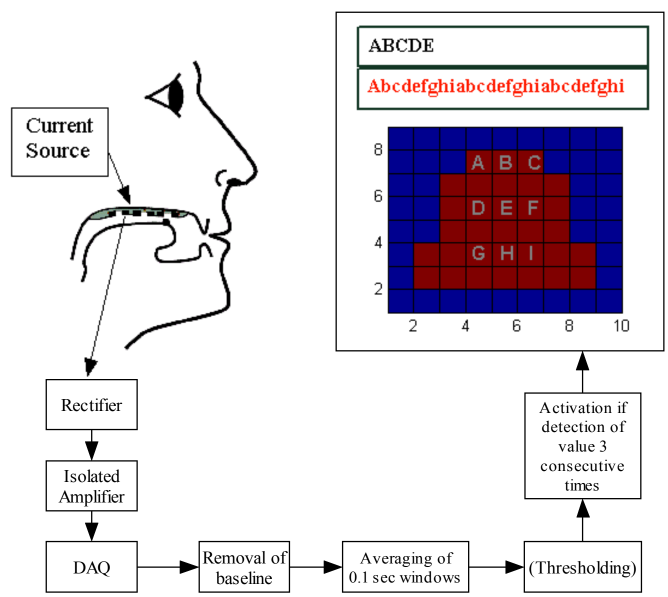

Figure 1. Signal processing for tongue-typing using the ITCI, from the sensor selection to the computer's visual display. Thresholding was used only for the first four subjects. From [13] with permission

TABLE I
TYPES OF SEQUENCES AND THE NUMBER TIMES TYPED OVER EACH TRAINING SESSION. TYPES 14 AND 15 ARE THE TEST SEQUENCES

Type	*Sequence*	*Times*
1	Aaaaaaaaaaaaaaaaaaaaaaaaaa	2
2	Bbbbbbbbbbbbbbbbbbbbbbbbb	2
...	...	...
9	Iiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii	2
10	Abcabcabcabcabcabcabcabcab	2
11	Defdefdefdefdefdefdefdefdefd	2
12	Ghighighighighighighighighi	2
13	Abcdefabcdefabcdefabcdef	2
14	Abcdefghiabcdefghiabcdefghi	9
15	Eafbgchdieafbgchdieafbgchdi	9
	Total	**44**

B. Measuring Performance

In Section II, two performance measures were defined: theoretical and experimental. Theoretical performance of a tongue-computer interface is hard to measure because the movement of the tongue in the palatal area is highly non-linear. Anterior and middle palatal areas are easier to access than posterior areas; therefore performance varies depending on where the sensor is located in the palatal. As Fitts' Law assumes a linear *ID* (2), mainly these linear models are used for finger, gaze or pen-based interfaces and there is no current model that involves tip-of-the tongue selection and most of the tongue movement models are developed for speech production. In order to measure theoretical performance, a non-linear version of Fitts' Law should be adapted to quantify tongue motor performance over a range of movement conditions around the palatal area.

For obtaining experimental performance, we measured input speed and accuracy on trial types 14 and 15 of our tongue-typing experiments using the ITCI (See table I) by adapting the *TP* rate introduced by Lewis for dictation performance [14]. It is defined as the number of words entered divided by the time taken to enter them added to the time taken to correct the errors. Using this model, (1) becomes:

$$TP_e = \frac{60}{5}\frac{corr+1}{t+err\ C} \tag{3}$$

Where 60/5 is the conversion factor for cps to wpm, *corr*+1 is the number of correct characters typed plus one, *err* is the number of incorrect characters typed, *t* is the total time (in seconds) of the trial, and *C* is the time spent for correcting one single error. As we do not correct throughout the test, we are dealing with insertion errors. Therefore, we adapt C as the average time to type one character on the corresponding trial.

The ITCI must map more than one character onto a key, therefore a method is needed to disambiguate between the possible character options when pressing a determined key, similar as it is done for typing with mobile phone keypads. To be able to compare ITCI performance with other text input interfaces performance, we are assuming that the disambiguation algorithm gives 1 keystroke/character (KSPC).

C. Learning Model

Theoretical performance may be very useful for optimizing the keyboard layout and also for setting upper bound limits. But it makes more sense to evaluate the learning process of tongue-typing using the experimental performance metric in (3). In order to quantify the effects of learning to type with the tongue, we made regression analyses to obtain fitted learning curves for TP_e. The power law of practice [9] tends to fit learning data from a variety of domains very well and is generally accepted as the function or law of motor learning. The model can be expressed as (4), where TP_n is throughput at the n^{th} trial, TP_1 is the estimated throughput for the 1st trial and *a* is the learning coefficient.

$$TP_n = TP_1 n^a \tag{4}$$

D. Data Analysis

As mentioned in section III-A, two different setups were used: shared channels for 4 subjects and independent channels for the other 4 subjects. These 2 setups are considered as between subject factors (A_i , i=[1,2]). All subjects trained for at least 3 sessions (U_j , j=[1,2,3]) and typed the same 15 sequence types (V_k , k=[1,…,15]). U_i and V_k are considered within subject factors. Each type of sequence was repeated 2 times for training sequences V_1 to V_{13} or 9 times for testing sequences V_{14} and V_{15}. These repetitions are considered as the repeated measures (X_l), where l=[1,2] or l=[1,…,9] for the training and testing sequences respectively. Three of the subjects performed one extra day of trials; this was done one week after the training experiments and is not considered for the statistical tests.

The TP data presents a skewed distribution, therefore to obtain statistical normality; the logarithmic variable transformation (5) was applied when performing mixed factorial repeated measures (RM) ANOVA: *A*x*U*x*V*x*X*. Otherwise, all TP_e values are reported without variable transformation.

$$TP_e' = Ln(TP_e + 1) \tag{5}$$

Each type of sequences was typed a different number of times depending on the type of sequence (see Table I). Therefore, RM ANOVA was performed separately for RMA_1=2x3x13x2 and for RMA_2=2x3x2x9, in order to avoid dealing with missing data.

IV. RESULTS

A. Differences in Performance

The Estimated Marginal means for RMA1 and RMA2 are shown in Figure 2 and Figure 3, respectively. There is a plot for each type of setup used in the experiments and the data is plotted for the different types of sequences and sessions. Logarithmic transformation is used as defined in (5) in order to obtain statistical normality.

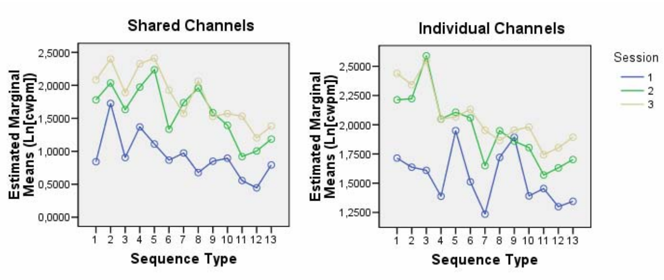

Figure 2. Performance marginal means for RMA_1 (sequence types 1-13) using logarithmic transformation. Marginal means are plotted for shared (left) and individual (right) channel setups depending on sequence type and session.

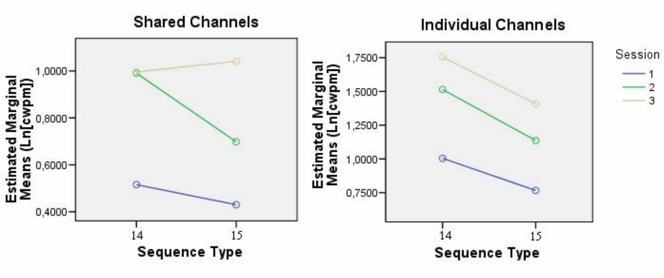

Figure 3. Performance marginal means for RMA_2 (sequence types 14-15) using logarithmic transformation. Marginal means are plotted for shared (left) and individual (right) channel setups depending on sequence type and session.

In all the plots a better performance when using individual channels (A_2) can be observed, specially for RMA_2, where even that the difference of performance was around the double compared to A_1, there was no significant main effect: $F(1,6) = 2.329,\ p > 0.05$. Therefore we analyzed the learning effects for all subjects together on the same charts.

It can be observed that performance improves significantly with each training session (U_j): $F(2,5) = 10.313,\ p < 0.02$. For RMA_1, there is a significant improvement from U_1 to U_2, but not from U_2 to U_3. In the case of RMA_2, there are even more significant effects of the training sessions: $F(2,5) = 19.125,\ p < 0.01$. Big improvement can be observed from U_1 to U_2 for both setups and sequence types. There is also a significant improvement from U_2 to U_3, with the exception of sequence type 14 when sharing channels.

Performance generally decreases with higher sequence type numbers (V_k). For RMA_1, there is no significant effect of the sequence type: $F(1,6) = 6.891,\ p > 0.2$, but for RMA_2, a significant effect of sequence type can be observed: $F(1,6) = 15.06,\ p < 0.01$. A noticeable decrease in performance from V_{14} to V_{15} was found, with the exception of sequence type 14 when sharing channels.

B. *Learning Effects*

Having the experimental performance results, we procede to evaluate the learning process of tongue-typing with regression analyses to obtain fitted learning curves for TP_e using the model presented in (4). In Table II we present the values of the learning parameters together with the squared error as a measure of goodness-of-fit. Subjects that present a large discrepancy between observed and expected values (R^2) were excluded for the plotting of the throughput data and learning curves in Figure 4. Including them would only insert noise in the plot, because such a poor fit cannot be trusted for model prediction.

TABLE II
LEARNING-CURVE EQUATION PARAMETERS FOR THE ESTIMATION OF TP_n

Subject	T_1	a	R^2	T_1	a	R^2
Shared Ch.		*Type 14*			*Type 15*	
Sub1*	0,191	0,505	0,236	0,287	0,419	0,363
Sub2*	0,651	0,338	0,214	0,270	0,564	0,621
Sub3*	0,303	0,877	0,699	0,236	0,803	0,741
Sub4	1,005	0,250	0,163	0,826	0,268	0,134
Individual Ch.		*Type 14*			*Type 15*	
Sub5	0,223	0,656	0,511	0,315	0,398	0,204
Sub6	0,617	0,602	0,492	0,451	0,515	0,619
Sub7	1,443	0,480	0,572	0,820	0,584	0,736
Sub8	1,041	0,802	0,696	1,692	0,385	0,769

* are subjects that performed a post-training session

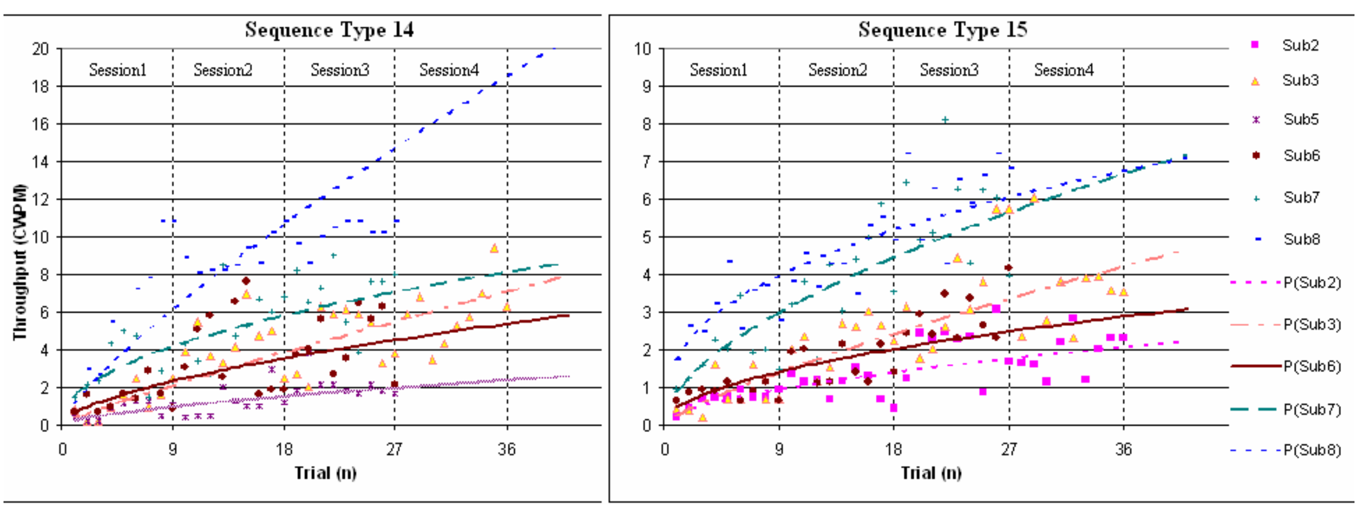

Figure 4. Learning curves for both test sequences (left: Type14, right: Type15) across 4 sessions for Sub2 and Sub3, and 3 sessions for the rest of the subjects. Learning curves with $R^2 < 0{,}3$ are not displayed. Throughput rates are presented in correct words per minute (CWPM) using (3).

Subjects 2 and 3 also performed one week post-training session. Analysis of this session revealed that the subjects continued to learn "off-line" and also showed improvement within the post-training exercise.

Type 15 is the most realistic sequence type for comparison with other text input interfaces because it consists on typing random character arrangements. If we continue extending the averaged learning curve for the sequence type 15, until expert performance levels (1000 trials), we can compare the typing rates with different hand-based and other alternative text input interfaces reported in literature (see Table III).

TABLE III
ESTIMATED ENTRY RATES FOR DIFFERENT TEXT INPUT INTERFACES AFTER 1000 TRIALS OR SENTENCES OF PRACTICE

Interface (Hand)	*TP(cwpm)*	*Alternative Interface*	*TP(cwpm)*
QWERTY[15]	150	*Dictation to ASR[15]*	107
Soft typing[15]	43	*Tongue(ITCI)*	25,36
Mobile(T9)[16]	45,7	*Head* [17]	12,1
Mobile(Letterwise)[16]	38,1	*Eye* [17]	9,36
Mobile(Multi-press)[16]	27,2	*Mouth-stick[1]*	8*
Mouse [17]	10,1	*Tongue (TTK)[1]*	4*

* are the results for the mean values over 9 sessions

V. DISCUSSION

To evaluate typing of real sentences, the ITCI must map more than one character onto a key, therefore an optimal character to key arrangement and grouping (layout) must be found. Also a method is needed to disambiguate between the possible character options when pressing a determined key. We assumed that the disambiguation algorithm gives 1 KSPC. However, a good disambiguation algorithm for mobile phones may give 1.15 KSPC [16], which would decrease our performance by approximately 13%.

Differences in performance attributed to sequence types are derived from the tongue's anatomy and the test difficulty. For V_1 to V_9, the user had to type one character repeatedly, for V_{10} to V_{12}, the user had to type one row repeatedly, V_{14} consisted in typing all the sensors in order and V_{15} was a randomized sequence. Repeated selection of the same sensor is much easier than selection of different sensors, also anterior and middle palatal areas are easier to access than posterior areas. Therefore, V_2 and V_5 are expected to present the highest performances, while V_{15} the lowest ones.

Acquiring the signals of each coil in separate channels improves the throughput rate means to more than the double for RMA_2. However, there is no statistical significance that tells us this improvement is due to the signal processing or to the subjects' performance. We may obtain significant difference if we involve more subjects on the experiments.

Regarding tongue-tasks motor learning, the results suggest that three consecutive 30-min tongue-selectivity training sessions result in an improvement of tongue-selectivity training skill one week post training. The ability of the tongue to quickly learn how to use the ITCI interfaces supports continued and improved development of the ITCI.

In general, tongue typing using the ITCI is slower than the standard QWERTY keyboard and other hand based input interfaces, but it outperforms other text input interfaces as assessed by data entry rates, such as eye, head and mouse interfaces. Therefore, the ITCI looks very promising as an alternative text input interface for physically disabled individuals.

ACKNOWLEDGMENT

The authors would like to acknowledge the valuable opinions and discussions of John P. Hansen, PhD. Also the support given by TKS A/S and the Danish Ministry of Science and Innovation.

REFERENCES

[1] C. Lau and S. O'Leary, "Comparison of computer interface devices for persons with severe physical disabilities," *The American journal of occupational therapy. : official publication of the American Occupational Therapy Association,* vol. 47, pp. 1022-1030, 1993.

[2] New Abilities Systems Inc, "Tongue Activated Communications Controller," US: WO 9307726, 1993.

[3] C. Clayton, R. G. S. Platts, M. Steinberg, and J. R. Hennequin, "Palatal tongue controller," *J. of Microcomputer Applications,* vol. 15, pp. 9-12, 1992.

[4] D. Kim, M. E. Tyler, and D. J. Beebe, "Development of a tongue-operated switch array as an alternative input device," *International Journal of Human-Computer Interaction,* vol. 18, pp. 19-38, 2005.

[5] L. N. S. Andreasen Struijk, "Tongue Based Control Method and System for Performing the Method," DK: WO 2006/105797, 2006.

[6] X. Huo, J. Wang, and M. Ghovanloo, "A magnetic wireless tongue-computer interface," in *Proceedings of the 3rd International IEEE EMBS Conference on Neural Engineering*, Kohala Coast, HI, 2007, pp. 322-326.

[7] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clinical Neurophysiology,* vol. 113, pp. 767-791, 2002.

[8] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement. 1954," *Journal of Experimental Psychology: General,* vol. 121, pp. 262-269, 1992.

[9] I. S. MacKenzie and R. W. Soukoreff, "Text entry for mobile computing: Models and methods, theory and practice," *Human-Computer Interaction,* vol. 17, pp. 147-198, 2002.

[10] S. Zhai, "Characterizing computer input with fitts' law parameters - The information and non-information aspects of pointing," *International Journal of Human Computer Studies,* vol. 61, pp. 791-809, 2004.

[11] D. W. Hansen, D. J. C. MacKay, J. P. Hansen, and M. Nielsen, "Eye tracking off the shelf," in *Proceedings of the 2004 symposium on Eye tracking research \& applications* San Antonio, Texas: ACM Press, 2004.

[12] P. M. Commarford and J. R. Lewis, "Models of throughput rates for dictation and voice spelling for handheld devices," *International Journal of Speech Technology,* vol. 7, pp. 69-79, 2004.

[13] L. N. S. Andreasen Struijk, "An inductive tongue computer interface for control of computers and assistive devices," *IEEE Transactions on Biomedical Engineering,* vol. 53, pp. 2594-2597, 2006.

[14] J. R. Lewis, "Effect of error correction strategy on speech dictation throughput," in *Proc. Human Factors & Ergonomics Soc. 43rd Annual Meeting*, 1999, pp. 457-461.

[15] R. K. Moore, "Modeling data entry rates for ASR and alternative input methods," in *INTERSPEECH* Jeju Island, Korea: ISCA Archive, 2004.

[16] I. S. MacKenzie, H. Kober, D. Smith, T. Jones, and E. Skepner, "LetterWise: Prefix-based disambiguation for mobile text input," in *UIST (User Interface Software and Technology): Proceedings of the ACM Symposium*, Orlando, FL, 2001, pp. 111-120.

[17] J. P. Hansen, T. Kristian, A. S. Johansen, K. Itoh, and H. Aoki, "Gaze typing compared with input by head and hand," in *Proceedings of the 2004 symposium on Eye tracking research \& applications* San Antonio, Texas: ACM Press, 2004.