MouthType: text entry by hand and mouth

Conference Paper · April 2004

DOI: 10.1145/985921.986070 · Source: CiteSeer

Michael J. Lyons
Ritsumeikan University
131 PUBLICATIONS **3,853** CITATIONS
SEE PROFILE

Chi-Ho Chan
University of Surrey
50 PUBLICATIONS **1,311** CITATIONS
SEE PROFILE

MouthType: Text Entry by Hand and Mouth

Michael J. Lyons **Chi-Ho Chan** **Nobuji Tetsutani**
ATR Intelligent Robotics and Communications Lab & Media Information Science Lab
2-2-2 Hikaridai, Keihanna Science City
Kyoto 619-0288 JAPAN
mlyons@atr.jp

Abstract

In this paper we describe a novel text entry method which uses coordinated motor action of hand and mouth. A vision based algorithm is used to gauge shape parameters of the cavity of the open mouth. These are mapped to a discrete set of input states which are combined with keypad input in a factorial manner to allow unambiguous input of a large number of symbols. The method is implemented and tested for an alphabetic writing system (the Roman alphabet) and a syllabic writing system (Japanese hiragana). We report the results of preliminary experiments to measure text entry speed and error rate.

Categories & Subject Descriptors:
H.5.2. User Interfaces: Input devices and strategies.

General Terms: Human Factors

Keywords:
Vision-based interface; mouth controller; mobile text entry.

INTRODUCTION

Writing is a fundamental technology that connects people in a way that stretches the temporal and spatial limitations of face-to-face interaction. Recent years have seen a tremendous surge in the numbers of writers and readers of text due to its widespread use in internet and mobile communications media.

Most text is created by manual action. However any human gesture can potentially be used for text entry. The role of the mouth in speech led us to consider the use of movements of the lower face for text creation. Automatic speech recognition and lip-reading are both active areas of research. However there is not yet a fully satisfactory speech-to-text system and robust, real-time automatic lip-reading is still a distant goal. Here we demonstrate the more modest concept of combining a small set of deliberate mouth gestures with manual action to enter text. One reason for doing this is that it enables single-keystroke text entry even with the small keyboards found on handheld devices. A further motivation is to create a text entry method for syllabic writing systems [2] without the unnatural requirement for alphabetic entry of the symbols of a syllabary.

RELATED WORK

A diversity of alternative methods has been proposed for mobile text entry. A review may be found in [3]. The closest related previous work is the recently proposed TiltText method [6], which uses orientation of a handset to disambiguate letters mapped to the phone keypad.

DESIGN & IMPLEMENTION

Vision System

We use a vision-based method to extract information about the shape of the open mouth. A lightweight headworn miniature camera, worn in a fashion similar to a headset microphone, captures an input image of the mouth (Figure 1). The shadow area of the mouth is segmented by selecting pixels obeying:

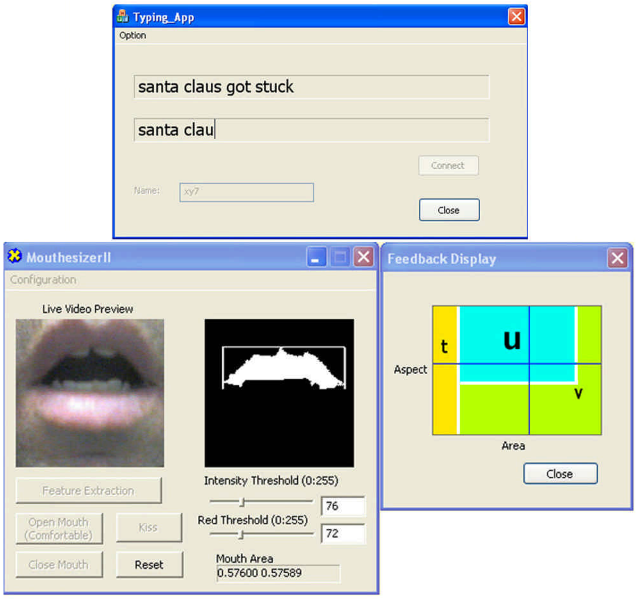
$$\mathbf{I} < \mathbf{I}_{\mathbf{max}} \text{ and } \mathbf{R} > \mathbf{R}_{\mathbf{min}}$$

I is the intensity of a colour pixel and **R** its red component. Intensity and red value thresholds are adjusted manually. We have also implemented a version which allows for adaptive thresholds. Further image processing selects the segmented region corresponding to the open mouth and eliminates noise due to image fluctuations and shadows. The area of the open mouth shadow is calculated as the total number of pixels in the segmented blob, while the aspect ratio is calculated as the height and width of the box bounding the blob. The algorithm runs at 30 frames/sec robustly under a range of lighting conditions. Comfortable maxima and minima of the area and aspect ratio parameters are input by the user in a calibration step. Recently our group has combined the system with a vision-based face-tracker, so that MouthType could also be used with a camera attached to a handheld device

Text Entry System

With MouthType, symbols are selected according an input mouth state ***M*** and a key press ***K.*** With *m* input states of the mouth and *k* keys, a total of $m \cdot k$ symbols may be mapped to the combinations (***M***, ***K***). We implemented prototypes for

CHI 2004, April 24–29, 2004, Vienna, Austria.
ACM 1-58113-703-6/04/0004.



santa claus got stuck

santa clau

Name: xy7

Intensity Threshold (0:255) 76

Red Threshold (0:255) 72

Mouth Area 0.57600 0.57589

Figure 1. English text entry with MouthType.

use with the Roman alphabet and the Japanese hiragana syllabary. A USB numeric keypad with the standard key mappings (Figure 2) for English and Japanese characters was used for manual input.

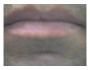
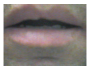
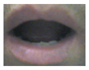
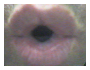
Alphabetic Text Entry: English

For alphabetic systems, the mouth shape to letter mapping is arbitrary. Four mouth shapes code letter order on the telephone keypad (Table 1). The first letter on a key (e.g. a, d, g) is selected by pressing that key while the mouth is closed. The second letter (b, e, h) is selected with a slightly open mouth, and the third (c, f, i) is selected with an open mouth. To enter 's' or 'z', the lips are puckered while pressing key '7' or '9'. Figure 1 illustrates MouthType text input of the letter 'u'. Pressing numeric key '8' selects the input map for the letters (t, u, v). Releasing the key selects the letter corresponding to the instantaneous area and aspect ratio of the mouth shadow. The input domain map display at the lower right of Figure 1 is optional: after a few trials one can input text without visual feedback of the area and aspect ratio parameters.

Syllabic Text Entry: Japanese

Japanese uses a mixture of three writing systems, one logographic, the kanji, and two homologous syllabaries, hiragana and katakana [2]. Hiragana can be used to enter symbols of all three systems. The basic hiragana syllabary is shown in Figure 2. Japanese syllable structure is fairly simple in that most syllables take the form CV (C = consonant, V= vowel), and there are only five vowels (**a, i, *u, e, o***). Most contemporary Japanese use the Roman alphabet to input the consonant and vowel of the kana. To enter the hiragana ***ra***, 'r' and then 'a' is pressed. Where appropriate, kana are converted to kanji via a selection menu, often using a predictive algorithm such as POBox [5]. Mappings of the entire kana exist for desktop keyboards but now are seldom used. On mobile phones, kana are input directly using MultiTap: repeatedly pressing a key cycles through the five vowel possibilities. An overview of Japanese text input is given in [5].

Table 1. Mouth shapes used for English text input.

Letter	**1**	**2**	**3**	**4**
Mouth Image				

	VOWEL					KEY
CONSONANT	あ a	い i	う u	え e	お o	1
	か ka	き ki	く ku	け ke	こ ko	2
	さ sa	し shi	す su	せ se	そ so	3
	た ta	ち chi	つ tsu	て te	と to	4
	な na	に ni	ぬ nu	ね ne	の no	5
	は ha	ひ hi	ふ hu	へ he	ほ ho	6
	ま ma	み mi	む mu	め me	も mo	7
	や ya		ゆ yu		よ yo	8
	ら ra	り ri	る ru	れ re	ろ ro	9
	わ wa		ん n		を wo	0

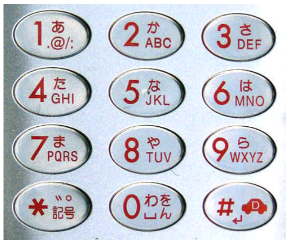
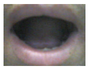
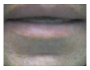
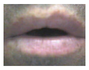
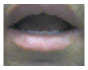
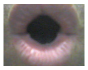
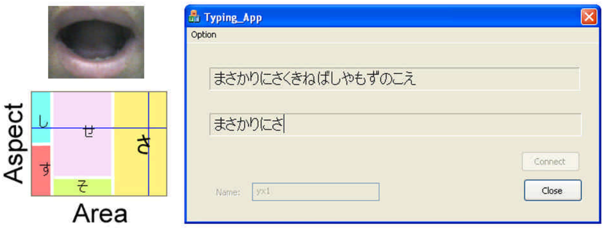


Figure 2 The basic Hiragana syllabary and a typical Japanese Mobile phone keypad.

The structure of the Japanese syllabary affords a phonetically derived MouthType input scheme. The mapping from the mouth area and aspect ratio to hiragana vowel categories was based on the shape of the mouth pronouncing the five vowels, as described semi-quantitatively in Table 2. A simple partition of the domain of the two input parameters satisfying the conditions in Table 2 is shown in Figure 3. In the input domain mapping diagram of figure 3, the aspect ratio of the mouth is greater

Table 2. Mouth shapes used for Japanese text input.

Vowel	**a**	**i**	**u**	**e**	**o**
Mouth Image					
Area	largest	small	small	mid-	mid-
Aspect Ratio	mid- or large	small	mid-	small or mid-	large



まさかりにさくきねばしやもずのこえ

まさかりにさ

Name: yx1

Figure 3. Japanese Hiragana entry with MouthType.

at the bottom of the vertical axis. This arrangement of the five vowels is similar to phonetic diagrams based on vowel formant frequencies or articulatory features [1].

Figure 3 shows a user entering the hiragana ***sa***: the user shapes the mouth as if pronouncing the vowel ***a*** while pressing the '3' key on the number keypad. ***Se*** is selected by shaping the mouth to the ***e*** vowel, while pressing the same key, whereas ***me*** is selected by pressing the '7' key with the same mouth shape.

A few additional keys are needed to code the complete hiragana. One key codes the diacritical marks for voiced (***ge***) or plosive (***pa***) consonants. Another is used to downshift to lowercase kana (e.g. small ***tsu*** or ***ya***). Kanji conversion was not implemented but could be added on using existing methods.

PRELIMINARY EVALUATION

We have not yet completed a full user study but report results of preliminary tests comparing MouthType with MultiTap for English and Japanese text entry. Subjects were the authors themselves. Each experiment consisted of a total of 10 sessions of 2 blocks with 8 phrases per block for each of the MultiTap and MouthType techniques. Entry technique alternated between blocks. Accurate text entry was enforced. Users were alerted of an error by a beep and had to delete the error and enter the correct character in order to proceed. Redundant key presses with MultiTap were not counted as errors. Key presses and releases were logged with timestamps.

English Text Entry

Experiment

Two subjects took part. Both have some prior experience with text entry using MultiTap but neither regularly sends text messages in English. A corpus of 500 short English phrases was used [4].

Results

The overall error rate was 3.1% for MouthType and 1.9% for MultiTap. MouthType error rates were higher for the letters in the interior of the input domain, which have greater exposure to transitions to the other two states. Error rates for the letters s and z were highest at 6.3% (Table 3). Averaged text entry rate development is shown in Figure 4. MouthType was faster than MultiTap for all sessions. The ratio of the number of key presses to complete the experiment for MultiTap compared to MouthType was 2.2.

Table 3. Error rate as a function of letter index.

Key Type	**Letter**			
	1	**2**	**3**	**4**
3-letter key	1.4%	3.8%	1.2%	-
4-letter key	0.4%	2.7%	1.2%	6.3%

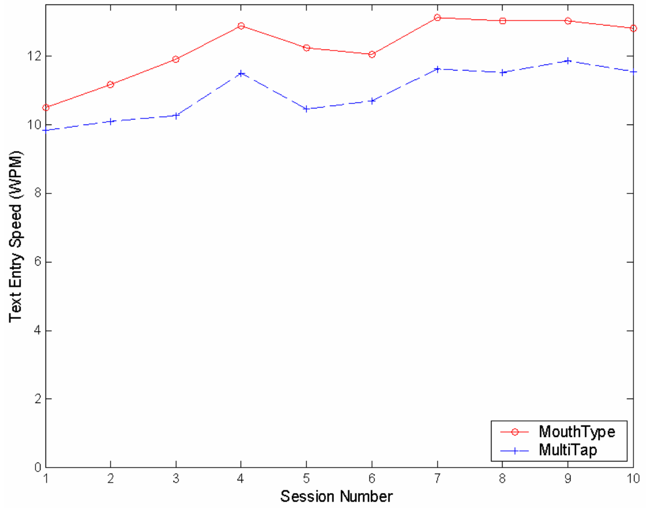


Figure 4. English text entry speed (wpm) vs. session.

Japanese Text Entry

Experiment

One subject took part. The subject is experienced in Japanese mobile phone text entry using MultiTap and POBox, composing about 5-10 Japanese messages per week. Each 17 kana phrase was chosen from a set of 80 haiku poems.

Results

The overall error rate was 8.7% for MouthType and 1.2% for MultiTap. Higher error rates were recorded for vowels in the interior of the input space (Table 4). Text entry rate development, in kana per minute (kpm) is shown in Figure 5. MouthType was considerably faster than MultiTap for all sessions, even taking into account the time required to correct the larger number of errors. With MouthType, text entry speed was still increasing at session 10, whereas with MultiTap there was little gain in performance over the entire experiment. The ratio of the number of key presses to complete the experiment for MultiTap compared to MouthType was 2.5.

Table 4. Error rate as a function of input vowel.

Vowel				
a	**i**	**u**	**e**	**o**
2.6%	7.9%	4.2%	12.9%	10.9%

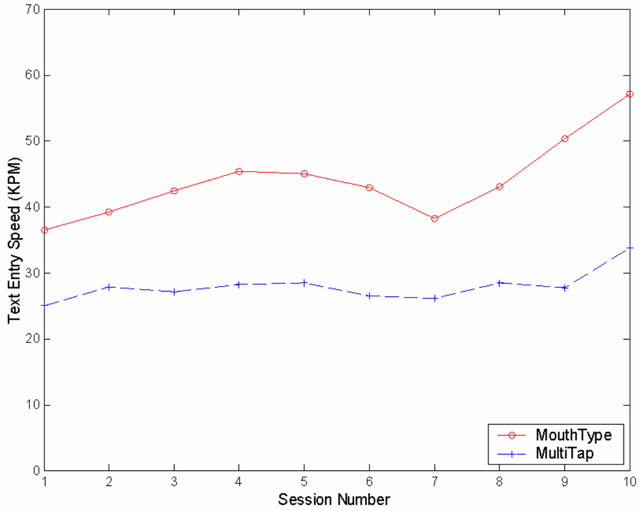


Figure 5. Hiragana text entry speed vs. session.

DISCUSSION

Learning to use MouthType involves the acquisition of two types of skill: (a) shaping the mouth to disambiguate letter or vowel, and (b) coordinating manual and mouth action. Neither skill was found to be particularly difficult to gain functional proficiency with: MouthType allows faster text entry than MultiTap both for English and Japanese text, as measured at constant accuracy of the input text. Skill (a) requires little learning for Japanese speakers using the Japanese version since the mouth shape to input character mapping is phonetically based. This may explain the large performance gain relative to MultiTap for Japanese text. The more modest performance gain for English text entry suggests that the arbitrary mouth shape to letter mapping presents greater difficulty than coordinating the two input streams, which is necessary for both Japanese and English text entry.

MouthType requires only 1 key stroke per character for both English and Japanese hiragana input, greatly reducing the load on the fingers. The motor action of the mouth required to operate the system is natural and did not result in fatigue during the course of the experiment. Text entry with MouthType could be considered as a mild form of exercise for the lower face.

MouthType has a much closer fit with the structure of Japanese writing system than either MultiTap, where one cycles through vowels with multiple presses, or input on QWERTY keyboards using the Roman alphabet, a method which has little intrinsic relation to Japanese script.

The same concept could be extended to text entry in other syllabic writing systems. With Inuktitut, a language spoken by the Inuit people of Nunavut, for example, consonants are represented using a small set of symbols with each of the three vowels (or absent vowel) indicated by four possible orientations of the symbol. It may also be feasible to adapt the concept to some of the scripts using syllabic alphabets, for which consistent modification of basic signs follows vowel changes. This includes many of the writing systems of South and Southeast Asia [2].

CONCLUSION

We have designed and implemented a prototype system which allows text entry by action of hand and mouth. The principle advantages of the system are that (a) it allows single keystroke text entry on small keyboards such as telephone keypads (b) it shifts part of the muscular load to an independent motor system closely affiliated with language and (c) for syllabic writing systems the system leverages existing user expertise, making the method easy to learn.

Results of a preliminary evaluation of text entry speed and accuracy are promising both for an alphabetic writing system (the Roman alphabet), and a syllabic writing system (the Japanese hiragana syllabary).

Future work will explore the following topics: the influence of input domain mapping on skill acquisition and error rates; an implementation using a camera on a handheld device; and application of the concept to other writing systems.

ACKNOWLEDGMENTS

The authors thank GC de Silva for helpful discussions. This work was supported in part by the Telecommunications Advancement Organization of Japan.

REFERENCES

[1] *Handbook of the International Phonetics Association*. Cambridge University Press, Cambridge UK, 1999.

[2] Daniels, P.T. and Bright, W. (eds.). *The World's Writing Systems*. Oxford University Press, New York, USA, 1996.

[3] MacKenzie, I.S. and Soukoroff, R.W. Text Entry for mobile computing: Models and methods, theory and practice. *Human-Computer Interaction 17*, 147 -198, 2002.

[4] MacKenzie, I.S. and Soukoroff, R.W. Phrase sets for evaluating text entry techniques. *Ext. Abstracts, CHI'2003*, 754-755.

[5] Masui, T. POBox: An efficient text input method for handheld and ubiquitous computers. *Proc HUC'99*, 289-300.

[6] Wigdor, D. and Balakrishnan, R. TiltText: Using tilt for text input to mobile phones. *Proc. UIST'2003*, CHI Letters 5(2), 81-90.