Natural Behavioral Patterns of Speech Recognition Error Recovery

Licheng Xue[1,2]
[1] State Key Laboratory of Brain and Cognition,
Institute of Psychology, Chinese Academy of Sciences
Beijing, P.R. China

Xianghong Sun[1]
[2] Graduate University of Chinese Academy of Sciences
Beijing, P.R. China

***Abstract*—Due to faster entry speed and lower attention demand, speech recognition is widely used in the multimodal text entry system. But error recovery is a big challenge to its usability for the accuracy of speech recognition is not good enough. However, few studies focused on the question how users recovered the errors in a multimodal entry system, especially in realistic scenarios. The present study investigated the natural behavioral patterns of speech recognition error recovery both in laboratory and field conditions. And further to examine whether there was any difference between these conditions. Results indicated that the handwriting modality was more frequently adopted in both laboratory and filed conditions. Similar behavioral patterns were found in these two conditions. And further, the usage of modality to recover error showed obvious individual variety in the laboratory condition, which was classified into three preference types. However, it was not apparently varied among individuals in the field condition. Furthermore, the entry modality had little effect on the choice of recovery modality. Finally, based on the findings in the present study, several applicable implications were provided.**

Keywords-multimodal text entry; speech recognition; error recovery; behavioral patterns

I. INTRODUCTION

Speech entry is extensively applied in the multimodal text entry system [1] [2], which has great advantages in terms of speed and attention demand [3] [4]. In a sense, speech entry provides an effective way to overcome difficulties of text entry for handheld communication devices [5] [6]. However, the accuracy of speech recognition is low because of the influence of various noises in the environment, which restricts the usability of the entire multimodal text entry system [7] [8] [9]. Therefore, the speech recognition error recovery is critical for the applicability of speech entry and even the multimodal entry system [10]. To our knowledge, few studies investigated the behavioral pattern of speech recognition error recovery in the multimodal text entry system, especially in a realistic condition [11]. Thus, this present study aimed to investigate the natural behavioral patterns as recovering the speech recognition error in laboratory and field conditions. The main purposes were to:

1. Investigate the behavior pattern of speech recognition error recovery in the laboratory test.

2. Explore the behavior pattern of speech recognition error recovery while walking.

3. Examine whether there was any difference in behavioral patterns between static and walking conditions.

4. Look into the influence of entry modality on the usage of error recovery modality.

II. EXPERIMENT 1: LABORATORY STUDY

A. Method

1) Participants

20 undergraduate and graduate students participated in the laboratory study. There were 10 males and 10 females with the average age of 23.3 ±2.3 years old. The history of using a mobile phone was 2 to 6 years. Most of them only used the keyboard; few had experience in using speech recognition system.

2) Material

Fourteen sentences were used in this experiment, without any numbers or English letters. There were four sentences in the "Speech entry-Free recovery" and "Handwriting entry-Free recovery" conditions respectively, and two sentences in each of other combinations (seen in Table1).

TABLE I. ENTRY-RECOVERY COMBINATIONS

	Entry Modality	Recovery Modality	Sentences
HS	Handwriting	Speech	2
SH	Speech	Handwriting	2
HH	Handwriting	Handwriting	2
S	Speech	Free to choose	4
H	Handwriting	Free to choose	4

Note: H – handwriting entry; S-speech entry

Handwriting recognition had a higher accuracy, so when entry modality was handwriting, participants were only allowed to select the first candidate character to ensure errors would emergence.

3) Equipment Interface

Nokia N810 mobile internet device (MID) was used as the experimental platform, and the interface was shown in Fig 1.

The first line on the interface displayed the text need to be input. Editing area was to show the input results and allowed test user to edit the result. The status bar signaled the status of input. The record button controlled the speech input. Once it

This study was supported by Nokia Research Center (NRC).

was pushed, the recording started, and the recording would not stop until the button was pushed again. Writing space allowed hand writing, and the candidate's area showed available words for user's choices. Click the Cancel button could clear the current input. In the function area, the modify button controlled the editing status with the edit space changing to be red. Delete meant to cancel one selected characters, and Clear meant to remove all the content in the edit space. Next button was used to update text for the new trial in the display space.

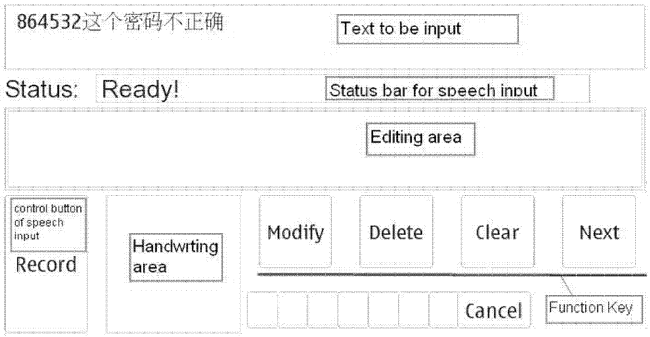

Figure 1. Program interface for tasks

4) Procedure

First of all, participants were required to input sentences with predefined modality, and then recover the recognition errors with modality required or been free to use any modality as shown in Table 1.

The first three combinations were randomly presented, to allow the participants to fully understand various kinds of recovery methods. And then the last two combinations were randomly presented to access to the real attitude when they could freely choose recovery modality.

Before the formal experiment, participants could practice to get familiar with the experimental procedure. And the results were recorded by the equipment of experimental platform.

B. Results

Only the results of the last two combinations were analyzed. The numbers of error in sentences, error recovered by speech and handwriting modalities were counted respectively.

The data were further adjusted in the calculation of the ratio of recovery modality. If the total number of errors was less than the total number of recoveries, added the difference into the total error count, so that it at least was equal to the entire recoveries. In addition, the default recovery efficiency of the handwriting was 1. That was, handwriting recovery could successfully recover one and only one error at one time:

The adjustment balanced the different efficiencies of the two error recovery methods.

The ratio of handwriting recovery = the number of handwriting recoveries/ number of errors adjusted (1)

The ratio of speech recovery = 1- the ratio of handwriting recovery (2)

1) Testing difference in the ratio of speech recognition error recovery between handwriting and speech modality

We used Wilcoxon signed-rank test to analyze the difference between these two kinds of recovery methods. As the result showed, the ratio of these two recovery methods significantly differed from each other (Z=2.76, p=0.006). Further, the ratio of handwriting recovery was larger than that of speech recovery, suggesting that participants preferred to use handwriting to recover errors.

2) The classification of the usage preference to error recovery modality

In order to further reveal the behavioral pattern in error recovery, the individual data was analyzed [12]. The participants were classified into three types according to the ratio of modality used (Table 2):

Type 1 Speech recovery preference: the ratio of speech recovery-the ratio of handwriting recovery > 30%
Type 2 Handwriting recovery preference: the ratio of handwriting recovery-the ratio of speech recovery > 30%
Type 3 Mixed type: the difference in ration between the handwriting and the speech recovery < 30%

TABLE II. MODALITY PREFERENCE OF USERS IN ERROR RECOVEY

Types	User ID
Speech preference	18, 14
Handwriting preference	6, 4, 11, 20, 7, 10, 2, 9, 3, 8, 15
Mixed type	1, 17, 19, 13, 12, 16, 5

In terms of the usage of error recovery modality, the handwriting recovery preference occupied the majority of participants, with 55%, mixed type followed, with 35%, and speech recovery preference was fewest, with 10%. Most of them with 85%, however, used both recovery methods, and a small number of them with 15% only selected the handwriting modality.

3) The effect of entry modality on the choice of the recovery modality

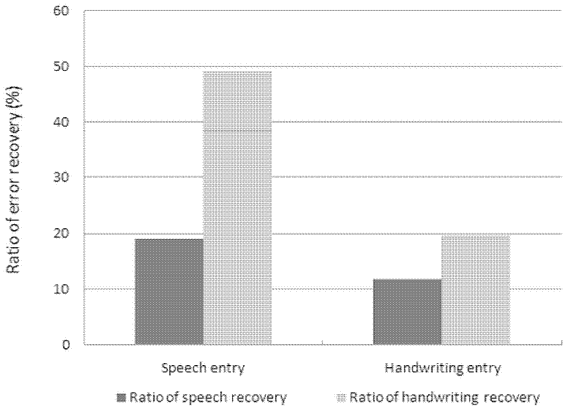

Figure 2. Ratio of error recovery modality usage in different entry modality

The result of Wilcoxon signed-rank test showed that in the speech entry condition, the ratio of handwriting recovery was significantly larger than that of speech recovery (Z=2.98, p=0.003). Similarly, in the handwriting entry condition, the ratio of handwriting recovery was significantly larger than that of speech recovery (Z=2.05, p=0.04). Overall, the choice of recovery modality was not obviously influenced by the entry method. Participants preferred to use handwriting recovery in both entry conditions.

C. Discussion

From the analysis above, we could see that the participants preferred to use handwriting recovery method. It was worth noting that, the majority of participants used both the speech and handwriting modalities to recover errors, while only a few always adopted handwriting recovery method.

Participants could be further classified into three types: speech recovery preference, handwriting recovery preference and mixed type. And the handwriting preference type occupied the majority, while the speech type was the fewest.

Entry method had little effect on the choice of recovery modality. In both handwriting and speech entry, participants preferred to use handwriting modality. Error recovery methods are relatively independent.

Taken altogether, more available error recovery modality will give more flexibility to the multimodal system. On one hand, the handwriting recovery was adopted frequently and widely, indicating that in editing text, the modality should be easy to access. On the other hand, the use of error recovery modality further showed inter-individual variation, including three main types. Most of participants belonged to handwriting recovery preference. Thus, when designing the text entry system, the individual difference should be carefully considered. It was necessary for an adaptive system, which should make proper changes according the individual differences.

III. EXPERIMENT 2: FILED STUDY

A. Method

1) Participants

Another ten undergraduate and graduate students took part in the field study (5 males and 5 females). Their average age was 24.1±2.7 years old. Most of them only used the keyboard entry method; few had experience in using speech recognition system.

2) Material and Equipment

The material and equipment were the same as those used in the laboratory study.

3) Procedure

Participants were asked to complete text entry task as walking in a quiet corridor, with suitable light. Before the experiment, they first practiced to use experimental equipment and 12 sentences were used to get familiar with the equipment and text entry task.

B. Results

The data analysis was the same as that used in the previous experiment.

1) Testing difference in the ratio of speech recognition error recovery between handwriting and speech modality

Wilcoxon signed-rank test was used to examine the difference between these two recovery methods. Results indicated that there was significant difference in the ratio between these two methods (Z=2.84, p=0.005), and the proportion of handwriting recovery (90.3%) was significantly larger than that of speech recovery.

2) The classification of the usage preference to error recovery modality

In order to further reveal the behavioral patterns in error recovery, we also analyzed the individual data (Table 3). The classification criteria were the same as Experiment 1.

Results in Table 3 showed that all of participants preferred to the handwriting recovery in walking, especially, half of them completely adopted handwriting modality.

TABLE III. MODALITY PREFERENCE OF USERS IN ERROR RECOVEY

Types	User ID
Speech preference	0
Handwriting preference	1, 2, 3, 4, 5, 6, 7, 8, 9, 10
Mixed type	0

3) The effect of entry method on the choice of the recovery modality

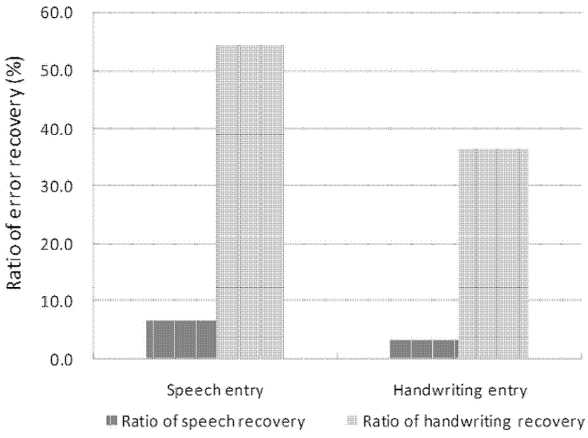

Figure 3. Ratio of error recovery modality usage in different entry modality

The results of Wilcoxon signed-rank test showed that the ratio of handwriting recovery was significantly larger than that of speech method in the speech entry (Z=2.67, p=0.008), while the ratio of handwriting recovery was also significantly larger than that of speech recovery in the handwriting entry (Z=2.80, p=0.005).

C. Discussion

These results in the field study indicated that under the walking condition, all participants preferred to use handwriting methods to recover error, and the choice of recovery modality was not significantly influenced by the entry method. Half of them adopted two methods and the other ones completely preferred to handwriting modality (reached 100%). This finding indicated that the usage of the modality to recover a speech recognition error was less varied among individuals as walking. And further, there may be difference between the laboratory and field conditions. Certain factors were found while walking, which may have important effect on the usage of modality to recover errors, such as the emergence of the pedestrians in the corridor, and the increased cognitive load during moving.

IV. CONCLUSION

Combined with the results in the laboratory and field experiments, we found that the handwriting modality was more frequently used for recovering speech recognition errors. There was no difference between two conditions in nature, but while walking, individual variety was less, and the usage of recovery modality was more hardly influenced by the entry methods.

Based on these findings in the study, several applicable implications were provide and should be paid more attention to. On one hand, give the priority to handwriting recovery method to meet the majority. On the other hand, the speech recovery method should be offered as an alternative to improve the adaptability of the system. The results could be an important reference for further system improvement.

REFERENCES

[1] Y. Y. Jiang, , X. G. Wang, F. Tian, X. Ao , G. Dai, and H. Wang, "Multimodal Chinese text entry with speech and keypad on mobile devices," Proceedings of the 13th international conference on Intelligent user interfaces, 2008, pp. 341-344.

[2] K. Kurihara, M. Goto, J. Ogata, and T. Igarashi, "Speech pen: predictive handwriting based on ambient multimodal recognition," Proceedings of the SIGCHI conference on human factors in computing systems, 2006, pp. 851-860.

[3] "Spoken multimodal human-computer dialogue in mobile environments," Series: Text, Speech and Language Technology , Vol. 28, W. Minker, D. Bühler , and L. Dybkj Eds, Kluwer Academic Pub, 2005, pp. 21-24.

[4] S. Oviatt, "Mutual disambiguation of recognition errors in a multimodel architecture," Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit, CHI' 99, 1999, pp. 576-583.

[5] S. A. Brewster, and P. G. Cryer, "Maximising screen-space on mobile computing devices," CHI'99 extended abstracts on Human factors in computing systems, 1999, pp. 224-225.

[6] R. L. Duchnicky , and P. A. Kolers, "Readability of text scrolled on visual display terminals as a function of window size," Human Factors, vol. 25, 1983, pp. 683.

[7] A. Alapetite, H. Andersen, and M. Hertzum, "Acceptance of speech recognition by physicians: A survey of expectations, experiences, and social influence," International Journal of Human-Computer Studies, vol. 67, 2009, pp. 36-49.

[8] C. M. Rebman Jr, "An exploratory study of the impact of training times on user acceptance of speech recognition systems and its managerial implications," International Journal of Innovation and Learning, vol. 3, 2006, pp. 607-61.

[9] H. Sacher, "Interactions in Chinese: designing interfaces for Asian languages," Interactions, vol. 5, 1998, pp. 28-38.

[10] J. Sturm, and L. Boves, "Effective error recovery strategies for multimodal form-filling applications," Speech Communication, vol. 45, 2005, pp. 289-303.

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

[12] S. Oviatt, , R. Coulston , and R. Lunsford, "Just Do What I Tell You: The Limited Impact of Instructions on Multimodal Integration Patterns," Lecture notes in computer science, 2005, pp. 261-270.