

Improving Human Interface in Drawing Tool Using Speech, Mouse and Key-board

Takuya Nishimoto, Nobutoshi Shida, Tetsunori Kobayashi, Katsuhiko Shirai

Department of Electrical Engineering, Waseda University
3-4-1 Okubo, Shinjuku-ku, Tokyo 169, Japan

Abstract

This paper focuses on the utility of speech input. We proposed some principles of human-computer interaction, which consist from the basic principles and organization principles of interface that are required for comfortable input systems. Then applying the principles, we discuss the desired organization of interface using speech, mouse and key-board and designed a multimodal drawing tool *S-tgif*. We also verify the benefits of speech input using the tool. As the results, the multimodal drawing tool with speech recognition reduces average operation time to 82%, the number of command inputs to 89%, and movement of mouse pointer to 53% in evaluation with inexperienced users. It is expected to make speech input more beneficial by applying our guidelines.

1 Introduction

Speech recognition technique had greatly improved in recent years, and many kinds of demonstration systems are realized in limited task domains[1][2]. Nevertheless, very few systems have shown the benefits of speech because the alternative methods such as keyboard, mouse, and other pointing devices are often more convenient for the task.

It comes from that the systematic investigation of human-machine interface has been neglected by researchers on speech recognition.

To realize useful spoken dialogue interface systems, it is important to investigate the principles of rapid and comfortable interface, and clarify the role of speech to satisfy the principles. For this purpose, the advantages of speech input to the other input methods must be discussed.

There are some attempts to clarify the characteristics of speech input compared to other input methods[3],[4],[5],[6],[1]. They are discussing the rapidness and error rates of speech input in comparison with other input methods such as key-board and mouse. Their result are very diverse and there is little agreement with the effectiveness of speech input.

Though most of their approaches assume that the speech input is the alternative way to the other input method, Martin[7] has worked from the viewpoint that speech input is one of multiple modalities which cooperatively enhances user's ability. He evaluated the utility of speech input in the context of VLSI CAD

and his result supported the benefits of speech input.

We also consider the effectiveness of speech input in multimodal input environment. We investigate the characteristics of speech input and the effective way of cooperation with other input methods from the viewpoint of *Principles of Interface*, in addition to regarding speech as an additional modality.

Many designers of interface proposed some guide for interface design[2],[8]. Most of their methods, however, don't help us finding what functions should be realized because their approaches are lacking fundamental rules. Even in Shneiderman's Eight Golden Rules, prescriptions and principles are not separated.

In this paper, we suggest the basic principles of interface and the organization principles of interface, classifying the basic requirements in interface systems and desired organizations that match various levels of users' skill[9]. We also investigated how speech input contributes to satisfy these principles, and developed a drawing system called *S-tgif* to examine the utility of speech input.

2 Basic Principles of Interface

From the viewpoint of human interface, we generally desire to use tools easily. This means that we wish to operate with less difficulty, that we wish to learn operations with less difficulty, and that we wish to know the permitted operations and consequences with ease.

We propose *the basic principles of interface*, what are concerned with the required functions for comfortable input system. We call these principles "Principles of less operation", "Principles of ease in learning operations" and "Principles of transparency of the system".

1. Principles of less operations:
 - (a) **Minimize the motion of position.** The movement of hands, finger and mouse pointer should be minimized.
 - (b) **Minimize the number of operations.** The number of operations for a command should be minimized.
2. Principles of ease in learning operations:

- (a) **Make easy to remind the command names.** Names of the command should be easy to remember. For the purpose, names should be associated with their functions and not be confusing with other commands.
- (b) **Keep consistency.** All commands should be operated with consistent methods within the application and across many applications.

3. Principles of transparency of the system:

- (a) **Make easy to perceive.** The status of the system and available operations should be understood easily.
- (b) **Make easy to estimate results.** The result produced by an operation should be always predictable.
- (c) **Offer perfect feedback.** Perfect feedback should be always given for the operational results.

3 Organization Principles of Interface

There are trade-off relations between basic principles.

Let us think of the operation with key-board for example. From the viewpoint of less operation, short command names are desirable and the best way of operation should be pressing one button of the key-board. However, for the purpose of learning commands easily, command names should be associated with the names of the functions. If we meet the demand, the methods of operation would require some meaningful words, such as names of commands in full spelling. This kind of trade-off relations can be found in many places.

The ratio of every principle's consideration depends on the experiences and frequency of use, and the matter of preference. To realize the input system that is widely accepted by various levels of users, the system should provide the best organization that satisfies each condition of users.

We propose *the organization principles of interface*, which guides how each basic principle should be considered in organizing system to meet the demand of every level of user.

1. Principle of supporting beginners:

Every input system for application should have common styles for those unfamiliar with the application or the computer system.

2. Principle of treating frequent users with priority:

Specialized usage of the application should be provided for those who use the system frequently.

3. Principle of suggesting beginners to use the skilled usage:

Every input system should have a guidance mechanism that indicates the skilled usage for beginners.

4 Multimodal Drawing Tool Using Speech

We design a multimodal drawing tool using speech, mouse and key-board.

In drawing task, selecting objects and directing positions on the canvas screen are often mixed with commands changing modes such as selecting, drawing rectangles, circles, texts. Therefore, operations are apt to be intricate with only mouse because pull-down or pop-up menus will let the mouse pointer go away from the position to which the operator is paying attention.

In this case, multimodal interface using mouse, key-board and speech input is expected to be effective. Applying the principles mentioned above, the following points have been considered in using speech input.

4.1 Applying Principles of Less Operations

- Changing modes with speech input is enabled, therefore the operator can use mouse only to draw pictures on canvas screen.
- Shortcut keys are assigned for one hand so the user can hold mouse with another hand.
- Shortcut by speech input enables to use more commands directly (with no tree menus) than the number of the buttons on key-board.

4.2 Applying Principles of Ease in Learning Operations

- The system will recognize all the words that associates with the functions. To delete an object, for example, the operator would remind many words such as "cut", "delete", "erase", "remove", etc. If the system can recognize only one word, the operator must learn what is the acceptable word, what is similar to learn peculiar operation. To make easy in learning operations, the system should recognize as many words that the user can remind as possible.

4.3 Applying Principles of Transparency of the System

- Speech recognition module always shows its status and recognition results, and the acceptance of words is also indicated with beep.
- The mode of the drawing system is shown with the figure of the mouse pointer so that the user can perceive the change of drawing mode without moving the gazing point.

4.4 Applying Organization Principles of Interface

- Speech input and other devices are available in any time so that the usage for beginners and skilled operator is permitted at the same time.
- To suggest to use abbreviations, shortcut keys related to the speech input is displayed on the screen.

5 Implementation

We develop a multimodal drawing system *S-tgif* (*Speech-input tgif*) based on a public domain software. Figure 1 is the operating style of the system. The operator inputs commands with speech, while holding mouse in one hand and key-board with the other hand.

The system has three modules. Figure 2 shows the windows of each modules.

5.1 Speech Recognition Server

The system has a speech recognition server that runs on a general-purpose workstation. The server communicates with the client application by the X-Window system protocols. It recognizes 76 isolated Japanese words using Hidden Markov Models. The calculations are processed in real-time on HP Series 735 workstation.

Available words list and recognized word is shown independently of the *tgif* application window, and the shortcut keys correspond to each speech command is shown together.

The system records the motion of mouse pointer and the time required for each evaluation. The numbers of operations done by each device are counted manually.

5.2 Rejection of Non-Command Utterances

In the situation that the speakers are allowed to utter whatever they think freely, they tend to include some talks to someone other than the recognizer while they talk to the recognizer. They might include some talks to themselves also. Such utterances not directed to the recognizer make it difficult for the system to recognize the utterances of users. So it is very important to distinguish the utterances directed to the system from those not directed to the system.

In *S-tgif*, we implement a prosody-based rejection mechanism, which uses the distributions of the logarithm power. It would be effective because the users tend to speak clearly to the system than to the others, and the direction of the speaker's face affect the power considerably.

6 Evaluation

Six subjects work on evaluation. All subjects are university students who uses computer everyday. They have no preliminary knowledge of *tgif* and *S-tgif*.

Evaluation is done by the following procedure.

First, we explain the outline of *tgif* and *S-tgif* and both usage of mouse and speech input. Then subjects practice on the system freely for fifteen minutes using speech, mouse and key-board.

The subjects are divided into two groups. In group A, subjects drew the figure X with speech in first trial and draw the figure Y without speech in second. In group B, they drew the figure X without speech in first trial and draw figure Y with speech in second trial.

In doing the tasks, precise size and layouts of the objects are not required, but some indications are to be obeyed about colors of some objects, size of fonts, and width of some lines.

After the task was finished, we ask some questions to survey the user-interface satisfaction.

We counted the required time to finish the task, the motion of mouse pointer and the number of command operations.

7 Results

Table 1 shows the ratio of devices that are used for commands.

Table 1: The ratio of devices which are used for commands(%)

	w/speech	w/o speech
speech input	91	0
mouse	9	100
key-board	0	0

The effect of using speech S is defined by following equation:

$$S = \sqrt{\frac{\alpha_1\beta_2}{\alpha_2\beta_1}} \quad (1)$$

where α_1 and α_2 denote the mean costs in the first and the second trial of group A respectively, β_1 and β_2 are those in the first and the second trial of group B. This equation is based on the assumption that the effect of using speech, the difference in two groups, the difference of the tasks, and the difference of the trial number affect on the costs independently.

As the results, speech inputs reduced the operation as shown by Table 4.

Error rate of speech recognition is 7%.

Table 4: Reduction of operations(%)

operation time	82
mouse movement	53
num. of commands	89

8 Discussion

In this system, speech input reduced the required operation time for a task, mouse pointer movement and the number of commands.

The results of the questionnaire (Table5) show the system with speech input is beneficial especially on the easiness of learning and reduction of difficulty. It can be concluded that the system showed the usability of speech input from the viewpoint of the basic principles of interface.

As Table 1 shows, most users used speech input though the mouse or key-board is also available. It should be noted that operations using key-board is difficult to learn. Even for experienced user, speech input would be convenience in small tasks like this.

In our observation, the number of cancellation commands used in the test session is very small in comparison with the number of speech recognition error. On



Figure 1: The operation style of S-tgif. The user input commands with speech, while operating key-board with left hand and mouse with right hand.

Table 2: Results of the experiments (group A)

1st with speech			2nd w/o speech		
Time (sec)	Movement (pixels)	Num of commands	Time (sec)	Movement (pixels)	Num of commands
463	17284	33	448	37435	33
608	23217	77	458	36257	56
420	20406	41	433	35418	35

Table 3: Results of the experiments (group B)

1st w/o speech			2nd with speech		
Time (sec)	Movement (pixels)	Num of commands	Time (sec)	Movement (pixels)	Num of commands
944	44364	83	345	23422	28
524	41630	54	321	18194	35
295	26207	27	395	14601	44

Table 5: Results of Questionnaire for Users' Satisfaction (Average, 3:neutral)

Rapidness of system response	3.8
Easiness of leaning usage	4.3
Easiness of knowing permitted operations	3.3
Easiness of handling objects on canvas	3.8
Reduction of difficulty	4.0

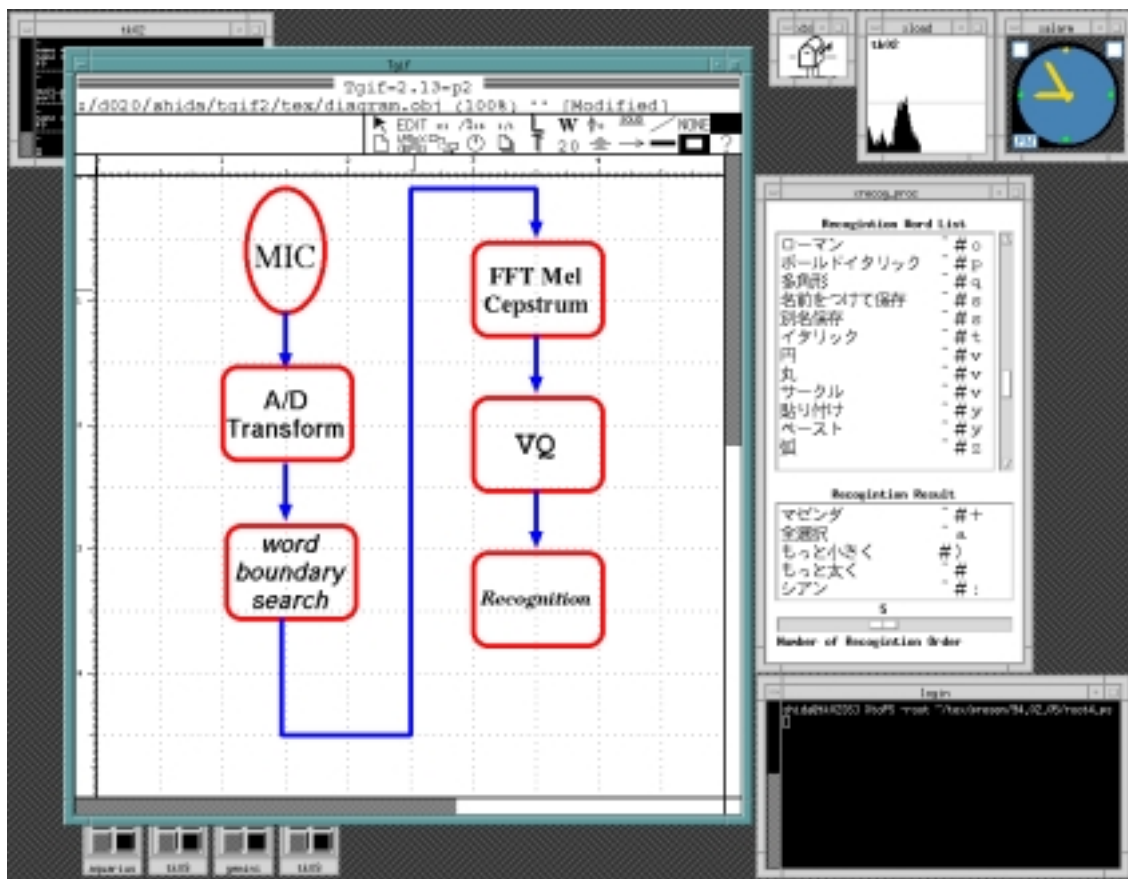


Figure 2: An example of the screen image of the system. Left window is the client system of the drawing module. Right window is the speech recognition server of the system.

some kind of situations in drawing application, recognition error can be recovered by inputting right command again without any undoing. This is supposed to be the reason that the number of cancellation was relatively small. It may be assumed that the drawing task is suitable for the application of speech input.

9 Conclusion

In this paper, we proposed the basic principles and organization principles from the viewpoint of interface usability. We investigated to apply the principles to drawing tool using speech input, then using the developed system, we verified the utility of speech input. Some future works are remained such as the investigation of the usage of speech output, speaker adaptation, customizing mechanism and the integration with gaze tracking system.

Acknowledgements

This study has been conducted as a part of the Project: Humanoid at HUREL(HUMANOID RESEARCH Laboratory), Advanced Research Center for Science and Engineering, Waseda University. The authors would like to thank ATR, NAMCO LIMITED, Nissan Motor Co.,Ltd, and YASKAWA ELECTRIC Corp., for their cooperation in this study.

References

- [1] Chris Schmandt, Debby Hindus, Mark S. Ackerman and Sanjay Manandher : "Observation on Using Speech Input for Window Navigation," in Human-Computer Interaction, D. Diaper et al. Edit, Elsevier Science Publishers B.V. (1990).
- [2] A.Rudnicky, J.Lunati and A.Franz: "Spoken language recognition in an office management domain," Proc. ICASSP'91,pp.829-832 (1991).
- [3] D.J.Cochran, M.W.Riley, L.A.Stewart : "An evaluation of the strengths, weaknesses and uses of voice input devices." Proceedings of the Human Factors Society - 24th Annual Meeting (1980).
- [4] J.M.Nye : "Human factors analysis of speech recognition systems," Speech Technology, 1, 50-57 (1982).
- [5] J.Leggett, G.Williams : "An empirical investigation of voice as an input modality for computer programming," International Journal of Man-Machine Studies, 21, 493-520 (1984).
- [6] R.Haller, H.Mutschler, M.Voss : "Comparison of input devices for correction of typing errors in office systems," Proceedings of INTERACT '84 (1984).
- [7] G.L.Martin : "The utility of speech input in user-computer interface," Int. J. Man-Machine Studies, 30, 355-375 (1989).
- [8] Ben Shneiderman: "Designing the User Interface," Second Edition, Addison-Wesley Publishing Company (1992).
- [9] T.Kobayashi, Y.Takeuchi and K.Shirai: "Multi-Modal Input Environment Using Speech, Mouse and Key-Board", Technical Report of IEICE, HC92-68 (1993).