



## IMPACT OF FIVE TASK-RELATED FACTORS ON THE CHOICE OF A VOCAL OR A MANUAL INPUT MODALITY

ROBERT, Jean-Marc, FISET, Jean-Yves, BERGERON, Gilles

Ecole Polytechnique de Montréal  
Département de génie industriel  
C. P. 6079, Succursale A  
Montréal, Québec  
H3C 3A7

### ABSTRACT

167 engineering students evaluated through a questionnaire the usefulness of a vocal and of a manual input modality for computerized systems. Five factors, grouped in triplets, were systematically manipulated in a set of scenarios: the type of task, the mode of presentation of information to the users, the task difficulty, the task duration, and the task frequency. Each triplet included the first two factors with one of the last three factors. The main results show that in "benign" conditions where the task is either short, not frequent, or easy, the users perceive either input modality as equally useful. In "stressful" conditions where the task is either long, frequent, or difficult, two general reactions are observed: the users prefer a manual input modality in spatial tasks where the information is presented visually, and slightly prefer a vocal input modality for all the other situations.

supports a compatibility principle whereby a verbal task is deemed to be compatible with an auditory stimulus and a speech response, while a spatial task is compatible with a visual stimulus and a manual response. The second factor manipulated in the scenarios is the mode of presentation of information to the users who perform the task: visual vs auditory. The two senses involved here are clearly responsible for most of the information that humans perceive. Furthermore, they are closely related to any tasks that can be performed with computerized systems and are involved in the compatibility principle mentioned above. These two basic factors were part of all the scenarios created for this study. The third factor was either task duration, task frequency, or task difficulty. These factors are sufficiently general to affect a large variety of tasks and were considered highly pertinent for voice applications.

### INTRODUCTION

Automatic Speech Recognition (ASR) is more and more appealing as an input mode for interacting with computerized systems. It offers major advantages for poor typists, handicapped, or people with their two hands busy at work; moreover, voice seems simply natural for some applications. Before adopting this new technology in the workplace, one must first know the performance of the ASR systems that are being considered, then evaluate their appropriateness for the specific applications wherein they will be used. Data on the performance of different ASR systems already exist (see [1] for a summary). Some data on the appropriateness of speech input for different applications also exist [2,3,4] but are still insufficient for predicting well how appropriate the systems will be.

The goal of this research is to measure the impact of five task-related factors on the choice of a vocal or a manual input modality for entering or manipulating information with a computerized system. Actually, subjects were asked to rate separately the usefulness of a vocal and of a manual input modality for each scenario, and their preference for an input modality was calculated afterwards. Five factors were systematically manipulated in these scenarios. The first factor is the type of tasks to be done with a computerized system: spatial vs verbal. These two types of tasks are borrowed from the Stimuli-Central Processing-Response model proposed by [2,5]. In this model, a task can be classified as verbal (or symbolic) if it requires "the use of language or other symbolic coding for its completion" while it is classified as spatial if it requires "a judgement or integration concerning the three axes of translation or orientation" [2]. This model also

### METHOD

**Experimental setting:** This setting consists of a 2x2x2 orthogonal plan which combines three task-related factors: (1) the type of task to be performed with a computerized system (spatial vs verbal), (2) the mode of presentation of information to the users (visual vs auditory), and (3) one of the following factors: a) the task difficulty [easy (little stress) vs difficult (task done with time pressure or in dangerous conditions)], b) the task duration [short (1 hour) vs long (all day)], and c) the task frequency [frequent (every day) vs not frequent (once a month)]. A pilot study conducted with a few subjects revealed that combining the five factors simultaneously generated too complex a situation. This option was therefore abandoned.

**Questionnaire:** The most practical way to collect data appeared to be the questionnaire. Simulation was considered impractical because of some of the factors that were tested (for example, it would be difficult to simulate a situation where a task occurs once a month). Data were collected by three sets of questionnaires in French, corresponding to the last three factors presented above. A questionnaire consisted of three parts. The first part allowed to collect subject's personal data regarding age, sex, training, computer use, typing skills, and personal deficiencies (if any) in vision, hearing, elocution, and manual dexterity. The second part presented a brief description of the vocal and of the manual input modalities for computerized systems. It also stated that the two input modalities had the same reliability and the same rapidity for entering and manipulating data, and were associated with the type of feedback (visual or auditory) deemed the most convenient to the subject. The

third part of the questionnaire included eight realistic scenarios which were presented in different orders for counterbalancing a possible order effect. A scenario combines a set of three factors presented above (see Appendix 1 for an example). For each scenario, the subject was asked to rate on a 5-point scale (1= not useful, 3= medium, 5= very useful) the usefulness of a vocal input modality, and on a similar scale the usefulness of a manual input modality. Subjects could also check "no opinion" on each scale.

**Subjects:** 167 engineering students from the Ecole Polytechnique de Montréal took part in the study (from the original 171 subjects, 4 were excluded because they had completed 50 % or less of the scales). They were split into three groups of 64, 44 and 59 subjects who respectively evaluated the three sets of questionnaires corresponding to task difficulty, task duration and task frequency. The subjects declared using a computer several hours a week (mean= 6.6; the range varies between 0 and 48 h/wt). When asked to evaluate their typing skills, 24 % of the subjects declared to be novice (0-15 wpm), 61 % intermediate (15-30 wpm), 14 % advanced (30-45 wpm), and 1 % very advanced (45 wpm and more). Some subjects declared having personal deficiencies in vision, hearing, elocution, or manual dexterity, but these were appropriately corrected. 77.2 % of the subjects were male and 22.8 % were female. Finally, the age varied between 20 and 47 years (mean =22.5).

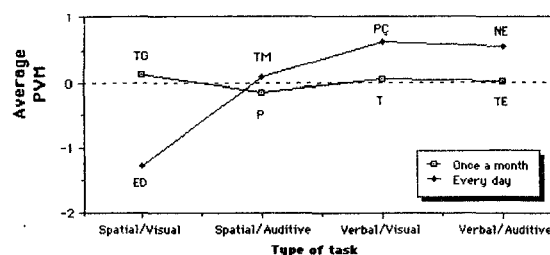
**Procedure:** The subjects were handed the questionnaire in the classroom. Data collection lasted between 10 and 15 minutes. Subjects were encouraged to write comments about the scenarios while completing the questionnaire.

## RESULTS

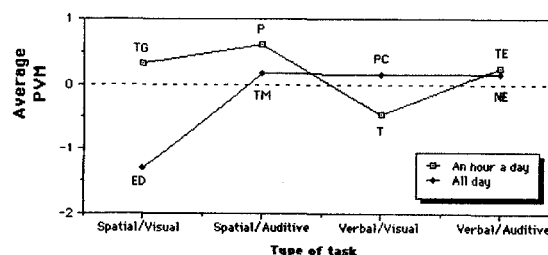
### Data compilation

Results are based on the compilation of the two ratings given by each subject for each scenario. The first rating bore upon the perceived usefulness of a vocal input modality (V) and the second, upon the perceived usefulness of a manual input modality (M). The second rating was subtracted from the first, yielding a composite measure of the preference of a vocal input modality over a manual one (PVM). Because each rating was made on a 5-point scale going from 1 to 5, the PVM measure can vary from -4 to +4 for each scenario. Adding the eight PVM measures for the eight scenarios evaluated by each subject yields an overall individual PVM measure, with a possible range of -32 to +32. The lowest overall individual PVM measure was -18 and the highest, +21. The distribution of these measures seems rather normal, with a mean of 1.11 and a standard deviation of 8.03. Therefore, on the average, the vocal input modality was considered slightly more useful than the manual input mode. However, the overall individual PVM measures vary considerably from one subject to another (see the large standard deviation). They also vary from one group of subjects to the other. In the group where the scenarios varied in terms of difficulty, the mean of the overall individual PVM measures was 2.89 (that is, 0.36 per scenario); in the group where they varied in terms of frequency, the mean of the overall individual PVM measures was 0.01; finally, in the group where the scenarios varied in terms of duration, the mean of the overall individual PVM measures was -0.02.

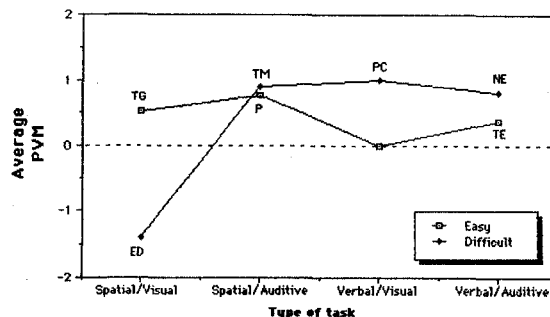
Surprisingly, the overall individual PVM measures are not related to the subjects' typing skills ( $F=1.24$ ;  $dl=3,163$ ;  $p > 0.25$ ).



Preference for input modality according to the type of task, mode of presentation and frequency  
Figure 1a



Preference for input modality according to the type of task, mode of presentation and duration  
Figure 1b



Preference for input modality according to the type of task, mode of presentation and difficulty  
Figure 1c

Legend:

ED: Engineering Drawing    NE: Number Entry  
PC: Process Control        P: Piloting  
T: Translation              TE: Text Entry  
TM: Telemanipulation      TG: Teleguiding

### Choice of an input modality

Figures 1a, 1b and 1c show the subjects' preference for an input modality according to the type of task, the mode of presentation, and either task frequency (1a), task duration (1b) or task difficulty (1c). Each point of a figure corresponds to the mean of the PVM measures for one scenario and all the subjects of a group; one can see the preference for a modality in each particular scenario.

Task frequency (Figure 1a): When the task is not frequent (i.e., performed once a month), the prospective users practically do not prefer any particular input

modality in any situation (mean PVMs = 0.12, -0.14, 0.05, 0.02). However, when the task is frequent (i.e., performed every day), three different outcomes may happen. When the task is spatial with the information presented visually, the users prefer a manual input modality (mean PVM = -1.26). When the task is spatial with information presented auditorily, the users consider the two input modalities as about equally useful (mean PVM = 0.09). Finally, when the tasks are verbal with information presented either visually or auditorily, the users tend to prefer a vocal input modality (mean PVM = 0.64, 0.60).

Task duration (Figure 1b): When the task is short (i.e., performed during an hour a day or less), the users tend to prefer a vocal input modality (mean PVMs = 0.33, 0.61, 0.25), except for the verbal/visual situations where they tend to prefer a manual input modality (mean PVM = -0.46). When the task is long (i.e., performed during all day), they prefer a manual input modality for spatial/visual situations (mean PVM = -1.30), but slightly prefer a vocal input modality for the three other types of situations (mean PVMs = 0.17, 0.14, 0.14).

Task difficulty (Figure 1c): When the task is described as easy, the users tend to prefer a vocal input modality (mean PVMs = 0.52, 0.77, 0.36), except for the verbal/visual situations where they are indifferent (mean PVM = 0.00). When the task is described as difficult, the users prefer a manual input modality for the spatial/visual situations (mean PVM = -1.41), but a vocal input modality for the 3 other situations (average PVMs = 0.91, 1.00, .81).

So, whenever a situation is presented as "benign" (i.e., either unfrequent, short, or easy) prospective users find the two input modalities as equally useful (with a small bias toward a vocal modality). On the other hand, when a situation is perceived as "stressful" (i.e., either frequent, long or difficult) the compatibility principle whereby the spatial/visual situations correspond to a manual modality is verified. However, the results concerning spatial/visual situations in stressful conditions, which strongly differ from the other results, deserve special attention. Two different scenarios -Engineering Drawing (ED) and Teleguiding (TG)- were used for testing the impact of frequency, duration, and difficulty. The ED scenario was systematically assigned to the "stressful" conditions, and the TG scenario, to the benign conditions. Then, it could happen that engineering drawing be responsible for the strong results in favor of a manual input modality. On the other hand, the results concerning spatial/visual situations in "benign" conditions, which do not differ significantly from the other results, could be due to the teleguiding task which would not be perceived as being as spatial as engineering drawing.

Overall, the fact that the subjects tend to find a vocal modality useful (except for the spatial/visual situations in "stressful" conditions) is consistent with the results of other studies (6,7).

#### CONCLUSION

This study has shown that the five factors manipulated herein strongly interact with each other and have an impact on the subjects preference for a vocal or manual input modality for computerized systems. This impact

goes with the compatibility principle whereby a spatial task with a visual mode of presentation is compatible with a manual modality, and a verbal task with an auditory mode of presentation is compatible with a speech modality.

A few promising directions for pursuing research on the appropriateness of voice input for different applications may be suggested. First, if the distinction between spatial and verbal tasks is being used, one should be careful and measure the degree to which tasks are perceived as spatial or verbal. These two adjectives could be the two ends of a continuum, and various tasks could be located at different points on this continuum. This might explain why two tasks, classified a priori, as spatial or verbal, might yield different ratings on a vocal or a manual scale, or yet be associated with different input modalities. Second, other factors such as safety and environment conditions (e.g., noise, dust, disturbances, etc.) could also be added to the five factors already manipulated in this study, in order to measure their impact on the choice of an input modality. Third, each factor could be examined separately in terms of impact on the choice of an input modality; experiments have to be designed accordingly so as to control the occurrence of a same factor in different scenarios.

#### REFERENCES

- [1] G L Martin, " The Utility of Speech Input in User-computer Interfaces", *International Journal of Man-Machines Studies*, Vol 30, pp 355-375: Apr, 1989
- [2] C D Wickens, D L Sandry, M Vidulich, " Compatibility and Resource Competition between Modalities of Input, Central Processing, and Output", *Human Factors*, Vol 25, no 2, pp 227-248: Apr 1983
- [3] Jones, D., Hapeshi, K., Franklin, C., "Design Guidelines for Speech Recognition Interfaces", *Applied Ergonomics*, Vol 20, pp 47-52: Mar, 1989
- [4] C A Simpson, M E McCauley, E F Roland, J C Ruth, B H Williges, " System Design for Speech Recognition and Generation", *Human Factors*, Vol 27 no 2, pp 115-141: Apr, 1985
- [5] C D Wickens, M Vidulich, D Sandra-Garza, "Principles of S-C-R Compatibility: The Role of Display-Control Location and Voice-Interactive Display-Control Interfacing", *Human Factors*, Vol 26, no 5, pp 533-543: Oct, 1985
- [6] C M Mitchell, M G Forren, " Multimodal User Input to Supervisory Control Systems: Voice-Augmented Keyboard", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-17, no. 4, pp 594-607: July/August 1987
- [7] J M Reising, D G Curry, " A comparison of voice and multifunction controls: logic design is the key", *Ergonomics*, Vol 30, no 7, pp 1063-1077: July, 1987

## APPENDIX 1

Example of a scenario  
(spatial task)

### ENGINEERING DRAWING

You are using a computer aided drawing (CAD) system to draw various mechanical parts. You are given a sketch of each part to be reproduced. To do so, you must use the CAD system to enter point coordinates, draw lines and geometric figures, rotate and translate the figures, change scale, etc. You execute this work **during all day**.

CIRCLE YOUR ANSWERS:

Usefulness of a SPEECH input mode:

1	2	3	4	5	X
---	---	---	---	---	---

not            medium            very            No  
useful            useful            useful            opinion

Usefulness of a MANUAL input mode:

1	2	3	4	5	X
---	---	---	---	---	---

not            medium            very            No  
useful            useful            useful            opinion