Discussing mathematics with my computer

Alistair D N Edwards [1], Evangelos Mitsopoulos [1], Eilis Montgomery [2]

1: Department of Computer Science, University of York, York, England YO10 5DD

2: Armature Limited, Enterprise House, 1 Apex View, Leeds, England, LS11 9BH

Abstract

Speech is the primary means of communication between people. It is the basis of two-way dialogues, whereby we speak and listen to what others say. Technology has just reached the stage whereby analogous bi-directional 'conversations' with machines are feasible, but the question remains open as to whether such dialogues are useful. We report on a study into the use of speech input and output in a particular application.

One motivation for using speech output from computers is as an alternative to visual presentation of information which is inaccessible to people who are blind. The *Maths* Project was concerned with overcoming the difficulties that blind people have studying mathematics because of its reliance on visual notations. A workstation was developed which presented mathematics in non-visual forms including synthetic speech. Input of control commands to the workstation was by voice or keyboard. The study compared keyboard input with speech.

The results suggest that speech input generally did *not* work well. There is a suggestion (based on an established model of cognition, ICS) that there is interference between the different parts of the task, the apprehension of spoken output, generation of spoken commands and the mental representation of the mathematics.

Introduction

The technologies of speech synthesis and speech recognition have reached such a stage that it is possible to create human-computer interfaces based on spoken dialogue. It is possible, but is it desirable? One of the worst reasons for using any technology is 'because we can'. One application in which there may be clear benefits of a conversational interface compared to conventional alternatives is for blind people. Visual, written forms of communication are essentially inaccessible to blind people, but spoken dialogue is not; blind people can take an equal part in spoken dialogues with other people, be they sighted or blind. The broad question addressed in this paper is whether interaction with a machine using two-way speech has advantages over other styles of interaction, though the specific application is somewhat narrower: mathematical problem solving.

Mathematics is essentially a cognitive activity, carried out largely 'in the head'. However, in practice, most mathematics relies on the use of visual notations, such as algebra. The precision and complexity of mathematics means that it usually has to be communicated in written forms; whenever two or more mathematicians gather together, there will be a blackboard or pencil and paper. Furthermore, even when working alone on a problem, a mathematician will rely on the *external memory* of a notepad to store intermediate results and the like. So, while there is no reason to believe that blind people should generally be less able mathematically, they are handicapped by the lack of an easily accessible external representation.

This was the foundation of the *Maths* Project[1]. Its aim was to provide a form of mathematical representation equivalent to a simple paper and pencil that could be used by someone with little or no sight. While it might seem superficially that this is a simple requirement, in practice it is rather challenging and the project was centred on the building of a multi-media computer workstation.

One of the output representations provided by the Workstation is speech. There are significant problems in using this form of representation, which mainly amount to avoiding ambiguity, but which were tackled in the Maths Project, and in its predecessor, the *Mathtalk* Project [1]. Another complication is due to the various forms of input required of the work station user: mathematical expression and control commands. It seems that the interaction might be simplified by allowing the use of speech input. This has the added attraction of having a system based on spoken dialogue: speech in and out. Attractive as this idea may be superficially, the question is left open as to whether it is practical, whether speech dialogues are practical and efficient. This paper reports an experiment which attempted to measure the effectiveness of speech input compared with a conventional ,keyboard-based interaction. Further details of the study can be found in[2] and [3].

1 *Maths* stands for Mathematical Access for TecHnology and Science and the project was funded by the European Commission (Tide Project number TP1033).

Experiment

A simple experiment was carried out comparing speech input to the Workstation with using a keyboard. Subjects were required to complete a set of simple mathematical exercises, using either speech input or keyboard commands. The commands controlled the 'browsing' of the mathematical expression, whereby the user could request the speaking of (for instance) the *next expression* or *previous term*. Commands were spoken simply by saying them (e.g. 'next expression'), while the keyboard commands were based on the initial letters of the commands with control codes. For instance, *next expression* was *Ctrl-n e*[2]. Speech input was handled by commercially available software (DragonDictate Classic version 1.0 for Windows) – which was programmed such that (for instance) when the user spoke the phrase 'next expression' (as a single utterance, with no pause between the words), the input codes *Ctrl-n e* were generated.

Subjects were sighted but had no access to visual representations. Sighted people were used mainly due to the difficulties in finding blind people who are suitably qualified mathematically [4].

It was not clear what differences could be expected in performance under the two conditions and therefore a wide set of data was

2 Ctrl-n is the code generated by holding down the *Control* key and pressing *n*.

collected. Many of the results did not show any apparent differences between the two conditions.(Note that there was no attempt to test the results for statistical significance, given that at five, the number of subjects was low). For instance, the number of mathematical errors and the time taken to complete the problems were approximately equal in both conditions. Similarly, when asked which input medium they preferred, approximately equal numbers of subjects indicated keyboard as speech.

Working on mathematics non-visually implies that the person must maintain mathematical information in their memory. For instance, in solving an equation, such as, $8 - 1 + 4 + 83 - 60 = x$, one has to maintain the sum of the left-hand side (7, 11, 94, 34). It seems that there are two ways of representing such information in memory, either as the sound of the number (repeated 'subvocally' in their head) or as an image of it written down.

Subjects were asked whether they heard or saw such results in their heads and this appeared to yield noticeable results. Their preferences as to speech or keyboard input and declared method of internal representation are seen to coincide, as in Table 1. However the apparent correlation is not what one might have expected. The verbal task of saying commands might be assumed to interfere with any internal verbal representation of the mathematics, and yet the results suggest the opposite.

It is possible to explain most of the results using an established model of human cognition, Interacting Cognitive Subsystems, ICS [5]. Space does not permit a full explanation here (see[3]), but the analysis leads to the conclusion that the finding that there was no preference of the speech input over keyboard input should not be surprising. One reason could have been that some subjects had good keyboard skills and found articulating commands in a robotic manner harder than locating the keys to be pressed (while rehearsing the intermediate result). For other subjects, the keyboard could have caused more disruption rather than speech. The relative difficulty between the two modes could also have been affected by the subject's ability to articulate commands in a clear and consistent way while rehearsing, as well as the subject's affective reaction to the frequent mis-recognitions. However, data are insufficient to allow for.

Subject	Preference		See	Hear
	Keyboard	Speech		
1	✓		✓	
2		✓		✓
3	✓		✓	
4		✓		✓
5	✓		✓	

Table 1. **Subjects' preferences for input medium and their stated method of remembering information.**

anything more than speculations.

The apparent preference for speech by people who hear the mathematics is harder to explain. Without further investigation, it is not possible to draw any secure conclusions. For instance, subjects were only asked once whether they saw or heard their internal representations, whereas it might have been that they shifted their strategy depending on the nature of the task (whether they were using speech or keyboard).

Conclusion

The conclusion is that speech input fails to demonstrate any significant advantage (in fact, fails to work well at all) for the very same reasons that made it feasible: technological advance. Until a powerful speech recognition system is feasible (with minimum response time and robust recognition) it is possible that in many circumstances traditional methods can be as good as or even better than speech input. However, if such technological limitations are overcome, it might be shown that discussing mathematics with a computer is superior to typing them.

Further investigation is warranted, possibly in different applications, to see whether similar results are repeated. Also, the ever-improving accuracy of speech recognition might yield some very different results.

References

[1] Stevens, R. (1996). Principles for the design of auditory interfaces to present complex information to blind computer users. DPhil Thesis, University of York, UK.

[2] Montgomery, E. (1998). *An evaluation of speech-based input to a non-visual multi-modal interface,* University of York, Department of Computer Science, MSc(IP) Project Report

[3] Edwards, A. D. N. Mitsopoulos, E. N and Montgomery, E. (1999)*Interaction based on spoken dialogues* (in preparation).

[4] Victoir, A., Linehan, C. and Stevens, R. (1997). *Final Evaluation Report,* Tide Initiative, Maths Project Deliverable No. D9.

[5] Barnard, P. and May, J. (1994). Interactions with Advanced Graphical Interfaces and the Deployment of Latent Human Knowledge. in F.Paterno (ed.) *Interactive Systems: Design, Specification and Verification.* Heidelberg: Springer-Verlag.

Acknowledgements

In addition to the University of York, the partners in the project were F H Papenmeier GmbH, Grif SA, Katholieke Universiteit Leuven and University College Cork, and their contribution is duly acknowledged.