E-learning Finds a Voice: a Study of a Speech-recognition Interface on an E-learning System.

S. PRITCHARD

Institute of Information & Mathematical Sciences
Massey University at Albany, Auckland, New Zealand
s.pritchard@massey.ac.nz

E-learning has been in use within commercial environments for many years although often under different names such as computer-based training. Since the early 1980s when they first began, the design of these systems has progressed with developments in technology and our understanding of how people learn. This paper presents a study into using a speech-recognition interface to e-learning, within the context of a commercial call centre. The aim is to investigate how such an interface affects the learning experience, and to develop a framework to guide speech-driven applications in the future.

---

1. Introduction

The basic premise of the research is that using speech in an interface to an e-learning system, both for command and control and in response to learning prompts, will produce a more effective and enjoyable learning experience.

The research questions that are being addressed are:
What is the impact of applying a multi-modal interface, including speech-recognition, to an e-learning system?
What framework can be produced to guide the application of such a multi-modal interface to an e-learning system?

In this paper we will consider the background in terms of e-learning and speech recognition, before going on to consider the context for the learning, call centres in New Zealand and in the UK, and preliminary results from a short study of call centres. The paper concludes with a discussion of the research strategy and conclusions.

2. Background

2.1. Technology-based training, or e-learning

Historically, technology-based training (TBT) covered a multitude of sins, and for some examples it certainly was a sin to expect anyone to enjoy learning from it. Going back to its roots in the early 1980s we find CAI (computer-aided instruction), CAL (computer-aided learning), CBT

(computer-based training), TET (technology-enabled training) and many other acronyms. What they all have in common is their objective of teaching a specified target audience about a particular subject area using a computer as the teaching medium. During the time from the 1980s until the present, the approach has moved from text-based page turning delivered on floppy disks, through various manifestations involving ever-improving graphics, interaction, sound, animations, simulation, moving images, user modelling and training management systems, delivered on floppy disks, CD-ROM, DVD, networks, intranet or Internet.

The phrase e-learning in this study means any type of system that has the objective of enabling its users to learn something, delivered by some form of electronic means, therefore in this research the phrases technology-based training and e-learning are used synonymously.

As Hills (1999) outlined, our learning potential increases with the number of learning channels that we are able to use. Up to now it has not been possible to access every learning channel as the "what we say" part of learning has not been available. This research seeks to investigate what the impact on learning is, of having that extra channel available, and how it should be used most effectively.

This seemingly simple statement encompasses issues in the technical area, how to make it work; in sociology and psychology, what is an appropriate way of using speech, for which types of learner and what types of learning problem; in interface design and usability, what is the most effective way of incorporating speech into the learning; and in evaluation, how do we measure the success of the undertaking?

Every e-learning designer strives to produce the most effective, interactive, stimulating and enjoyable learning experience possible, however there are limiting factors in our knowledge of effective design and the technical feasibility of achieving it.

In this research we will investigate how some of the limitations can be overcome by the theoretical and empirical understanding of applying speech-recognition technology to the e-learning interface.

2.2. Speech recognition

There are two types of speech recognition: trained speech recognition involves a specified user training the software to recognise his/her own voice; independent speech recognition software uses different types of algorithms to recognise speech from any user. In the context of an e-learning system it is independent speech recognition that is needed; it is not acceptable or appropriate to ask each new trainee to spend half-an-hour training the software before undertaking the training.

3. The Context

There are several models for delivery of e-learning in the commercial environment. The one that potentially accrues the greatest benefit to the organisation is delivery at the users' own workstations using an intranet. This gives all of the advantages:

- Minimum possible time away from the job
- No travel or accommodation costs associated with off-site training courses
- Updating of learning content happens reliable and on time at all locations (particularly important for product knowledge, such as for new initiatives or promotions)
- Training can be taken in small, manageable chunks to increase the likelihood of retention
- Training can be taken on a just-in-time basis, at the time that a procedure is about to be performed, or a skill-set used
- Training management programs can closely monitor the progress of each employee.

(Faint 1998; Horseman 1998)

If this "at the work-station" model is in use, interacting with the training using voice may not be acceptable either from the learner's view-point or for other people in close proximity, if the environment is particularly quiet. One of the advantages of using a computer as a learning medium is the privacy it affords less confident learners. Where responses to a training program are given by voice, the private nature of the activity is compromised.

With this in mind, call centres were chosen as the context of the study. Call centres are generally highly technology dependent, so are most likely to be using technology-based training, and due to the nature of the job, voice would be an acceptable and unobtrusive interaction mode. Call centres have a high staff turnover (Sayers et al,2003) resulting in a significant training requirement, and therefore a good supply of potential test subjects for trialling and evaluation purposes.

In Auckland, contacts were made through TUANZ (The Telecom Users Association of NZ) and to date, three interviews have been completed in call centres around Auckland.

Two interviews were completed during a recent research visit to the UK, with the potential of two more telephone interviews to be arranged.

All of the interviews were recorded and followed a semi-structured interview technique. These interviews are currently being analysed; preliminary results are discussed below.

4. The call centre study

Whilst the study is incomplete and the interviews conducted so far are still in the transcription process, there are some differences that can be seen.

There was a marked difference between the attitude of New Zealand and UK call centres. The New Zealand contacts were interested in any initiative that may improve their own training regimes and very willing to give time to talk about it. However, these contacts were the ones who had responded to an invitation to become involved, a positive attitude therefore is to be expected, and finding an hour to spare during the day for an interview did not appear to be a problem.

In the UK the situation, and the response, was considerably different. With no equivalent to TUANZ available it was a cold-call situation, so it is hardly surprising that there was only one positive response amongst twelve contacts, but in almost every case, the reason given for not getting involved was being far too busy to spend the time. The impression was one of far more frenetic activity and more stress amongst the contact centre managers and trainers. The second UK interview, resulting not from a cold-call but from a previous contact, was also extremely difficult to arrange due to the time pressures that the subject was under.

For both UK contacts, the centre was in the process of introducing new technology that would ensure that incoming calls could be answered by any agent, irrespective of geographical position, and that all agents have access to all customer databases. This requirement was due to takeovers and reorganisation within the companies. Both of the New Zealand companies were far more stable in their technology.

An interesting difference between the two countries is in the size of the centres, and consequently their training requirement. The UK centres investigated were very large; one of the subjects was a major utility company that expected to have to train between 1500 and 3000 new recruits annually, has to employ agency staff in order to cover their call volumes, thus exacerbating their training problems due to the temporary nature of many agency staff, and had to close down a call centre that was in a West London suburb due to the impossibility of recruiting enough staff in that area. In contrast, the New Zealand centres were on a smaller scale with lower staff turnover in general and a consequentially reduced training burden. This could be expected to have an effect on the uptake of new technology such as voice recognition, in training systems.

In two of the three NZ centres studied, one bank and one IT based company, the findings of Sayers et al (2003) is echoed in that there is a well established career path for call centre representatives to move into other areas of the business, and due to the extensive training received as

a customer service representative in the call centre, such employees are welcomed and tend to do outstandingly well in their new roles. This ability to progress within the company means that, whilst staff turnover within the call centre remains high, a significant percentage of those remain within the organisation, thus retaining the knowledge and skills learnt.

The ability to provide this career path appears to depend both on the nature of the employing organisation and its attitude to staff; opportunities within the organisation need to exist, and a willingness to support staff in their effort to transfer. It was noticeable that the UK centres laid greatest emphasis on the need to retain staff within the call centre in order to deliver the service levels required.

In terms of learning environment all of the call centres studied had a similar approach to the initial training period for new recruits. There was a six to twelve week classroom-based induction learning about the company, its products and services, its computer systems, the call centre technology and the procedures required for the job. Time "on the floor" is included in this period, when an extra plug-in headset is used so that the trainees can listen-in to calls being taken by experienced staff, and observe. The time at which trainees start to take calls differs across the companies, but all recruits are closely supervised in the initial period either in a one-to-one 'buddy' system using the extra headset as before, or in a closely supervised crèche arrangement.

The only centre that was trying to use e-learning during the initial training period was, understandably, the UK utility that had the largest training need in terms of numbers per annum. Here, a recent experience with poor quality e-learning has meant back-tracking on an original plan for unsupported e-learning delivered at the workstation, in favour of trainer-supported e-learning delivered in a separate training room. This change was in response to the trainees' reaction to the original arrangement; basically they refused to cooperate with it – a clear indication of the importance of gaining user acceptance for an e-learning implementation.

5. Discussion

The area of speech recognition has moved very quickly in the last few years (Juang and Furui 2000) and it is reasonable to expect it to continue to advance significantly in the next few years.

The first stage of the research, currently underway, is to investigate the context area of the research, call centres, and in doing so, to establish future test beds for the evaluation of resulting systems.

The literature review will need to be extensive covering the fields of education, AI, psychology, interface design, usability and evaluation. The need to keep the review up-to-date throughout the duration of the research is even more pressing in a field like this where progress and change is happening so fast. From the review a theoretical hypothesis will be developed as to the effect of a speech-driven interface on a learning experience.

The next stage is to identify a suitable piece of e-learning that is in constant use within a call centre, and add a voice interface in addition to the existing mouse/keyboard interface. This will enable the resulting multi-modal interface to be trialled against the original version in the call centre, using the trainees split randomly across the two versions. The assessment will be both quantitative and qualitative, user satisfaction being at least as important as learning outcomes.
This stage will establish technical approaches and techniques for adding a voice interface, as well as addressing design issues. It is highly likely that this will be an iterative process similar to the study undertaken by Yankelovitch, Levow and Marx (1995) who tested each iteration of interface design with a new cohort of users.

This will be the first opportunity to test the theoretical hypothesis and to start to develop a framework to guide future applications.

After the successful completion of this initial phase, an attempt will be made to introduce some level of machine understanding of what has been said. This is expected to enable new approaches to the presentation of the learning content itself. Currently, a piece of standard e-learning is likely to teach two or three learning points, then reinforce the learning using multiple-choice questions. Multiple-choice questions are used due to the difficulty of parsing free-format short answers for meaning, and because to request typed answers assumes keyboard skills that, in many cases, cannot be guaranteed.

Using a speech interface removes the constraint of keyboard skills and allows a dialogue approach to be developed.

Ongoing research can address the questions of what differences are apparent in different contexts, and how interfaces should be modified for different task characteristics or user characteristics.

6. Conclusions

The aim of this study is to investigate the effects of using a speech interface on a piece of e-learning within the setting of a call centre, and to establish a framework to guide the construction of such interfaces. At the conclusion of the study the framework will benefit training

producers and help to provide the most effective training systems possible for this volatile environment.

Further areas of research to translate this framework into other commercial environments would be a move towards generalising the results for use in other learning situations.

The call centres studied were in unanimous agreement as to the "holy grail" of machine-based training using speech; a system that trainees can interact with using natural language that can simulate an incoming call, thus enabling them to practise their skills without tying up valuable trainer resource on a one-to-one basis.

Whilst accepting that this may be their ideal, there are many smaller steps to be taken before that can become a reality.

References

Faint R (1998) Intranets for learners. *IT Skills* July p30

Hills H (1999). *What works well in the management of TBT*. Presentation to the association for computers in Training, Sheffield, March 17, p12.

Horseman C (1998) *Intranets: the secrets trainers should know*. Presentation to Training Solutions Conference. July. NEC

Juang, B.H. and Furui, S. (2000). Automatic recognition and understanding of spoken language – a first step toward natural human-machine communication. *Proceedings of the IEEE* 88(8): 1142-1165

Sayers J, Barney A, Page C, Naidoo K, (2003). A provisional "thumbs up" to New Zealand bank call centres. *Business Review*, 5 (1), Auckland, University of Auckland Business School.

Yankelovich, Nicole, Levow, Gina-Anne, Marx, Matt, (1995) Designing Speech Acts: Issues in speech user interfaces. *CHI95 Proceedings*