Experiences Collecting Genuine Spoken Enquiries using WOZ Techniques

Roger Moore and Angela Morris

Speech Research Unit
Defence Research Agency Malvern (RSRE)
St. Andrews Road, Malvern, Worcs., WR14 3PS, United Kingdom

ABSTRACT

This paper presents SRU's first experiences collecting unscripted speech data using the 'Wizard of Oz' technique to provide a *genuine* telephone-based route planning service. Although only a limited quantity of data has been collected so far, several valuable insights into the nature of future speech-based human-machine interaction have been obtained.

1. INTRODUCTION

Many laboratories have now used the so-called 'Wizard of Oz' (WOZ) technique for eliciting spontaneous spoken human-machine dialogue in order (i) to study the resulting speech and (ii) to evaluate the necessary speech technology and natural language processing systems [1]. The technique is particularly valuable because it enables user behaviour to be studied under conditions which are not constrained by the limitations of current technological (or theoretical) capabilities.

However, many such exercises involve 'volunteer' users whose behaviour is prescribed by a pre-prepared scenario [6] thereby removing one potentially crucial aspect of human-machine interaction, namely any behavioural events which are unique to *genuine* (i.e. motivated) and un-preprepared transactions [4].

This paper presents SRU's first experiences collecting unscripted speech data using the WOZ technique by the provision of a *genuine* voice-based telephone enquiry service to personnel on the RSRE site.

2. THE TASK DOMAIN

The enquiry service was configured around a commercially available route planning software package. The package runs on a PC and contains map and gazetteer information covering the majority of the United Kingdom. Its main feature is its ability to find the shortest and/or quickest routes between two locations in accordance with a range of specifyable variables such as preferences for certain classes of roads and driving speeds. Alternative routes can also be found.

3. THE WIZARD

Clearly the behaviour of the wizard will very much influence the nature of the resulting corpus, and constraints (such as restricting the vocabulary) can be placed on a wizard in a variety of ways [2].

However, in order not to restrict or influence callers' behaviour, it was decided that very few restrictions should be placed on the wizard apart from the use of a stock opening phrase and the provision of a few standard reply templates simply in order to reduce the wizard's work load.

In general, the design for the wizard's behaviour was based on information derived from the procedures employed by a commercial company who already provide a route planning service over the telephone (in this case the enquiries being made by tone dialling) based on the same software package.

Every attempt was made to remove all distinctly human characteristics from the wizard's speech such as false starts and stutters, and great care was taken to ensure that breath noise and key clicks were not audible to the caller.

4. THE EXPERIMENTAL CONFIGURATION

As well as implementing a genuine telephone-based enquiry service using WOZ techniques, it was also decided to compare wizard-type transactions with normal human-human interaction for the same task. Hence the experimental set-up was configured to operate with two incoming telephone lines - one assigned to the normal human operator and one assigned to the wizard. Appropriate equipment was installed to provide automatic detection of incoming calls and initiation of recording and digitisation.

In order for there to be minimal differences between the operator's behaviour in both the human-human and human-wizard conditions, the same operator was used in each case. As a consequence the only difference between

the two conditions was that the wizard's natural voice was modified using a 'voice disguise' unit. This device changed the talker's pitch and then combined the natural and altered signals to produce a highly synchronised duet effect. It was found that this provided a voice which was unnatural (indeed 'robotic') and yet fully intelligible [5].

The latter point was considered very important since it was anticipated that the quality of the wizard's voice would affect the user's perception of the system's capabilities; high voice quality being likely to suggest a system of high capabilities while a low voice quality would not only imply a system of poor capabilities but might lead to excessive confirmatory dialogue if the user had difficulty understanding the response [2].

The service was made available on each line for alternate half day sessions from 10 a.m. to 12 a.m. and 2 p.m. to 4 p.m. Whilst one number was on-line the other was connected to an answering machine which requested the caller to try the alternative number.

The table of directions that the route planning software package produces is an essential part of the information service. To overcome the lack of screen display with a telephone-based service it was decided that the printed table could be sent to the caller through the internal mail system (this also served to ensure the identity and location of the caller was known in order to send a questionnaire at a later stage). Callers were also given the option of having the route information read out during the call.
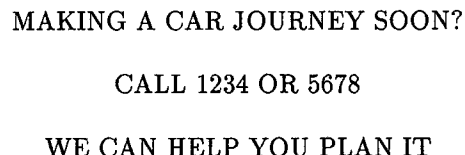
The route planning package was configured to show no particular preference for road type and the road speeds were set at the national speed limits.

5. INSTRUCTIONS TO THE CALLERS

A poster advertising the service was circulated for display on site noticeboards. Also an electronic advertisement was placed in the central computing facility.

Since the emphasis of the exercise was to collect data from genuine enquiries, the advertisments made no mention of either the Speech Research Unit or of the difference between the two available telephone numbers, nor did they specify that the service was experimental, computer-based or automatic in any way.

On receipt of a call, the operator (in human or wizard mode) always used the following introductory announcement:- "Welcome to the route planning service - how can I help you?"

```
MAKING A CAR JOURNEY SOON?

CALL 1234 OR 5678

WE CAN HELP YOU PLAN IT
```

Figure 1: The advertisement for the service

6. RESULTS

The service went 'live' for the first time during November 1991 but, due to the limited advertising campaign and a relatively small population of potential users, the number of calls received (by the end of December 1991) was rather smaller than had been initially hoped for. Nevertheless, the data collected during that period already capture some interesting general features of *genuine* spoken human-machine interaction.

For example, with some calls there was considerable background office noise (some callers appeared to be using loud-speaking telephones). Also, callers occasionally chuckled to themselves or made asides to other people in their vicinity (including statements along the lines of "Hey, I'm talking to a machine" and constant references to 'it') - although this confirmed that the callers were convinced by the wizard's voice it also indicated that they believed that the system automatically knew when it was being addressed! Some callers interrupted the wizard, and at least one mimicked the robotic style of the wizard's voice.

It was also noticeable that, although the human-wizard dialogues were all concerned with planning particular routes, most of the human-human dialogues were about the nature of the service itself. In other words, the users who dealt with the wizard seemed to assume that such a system would not be able to provide explanations about what it could and couldn't do - and so they didn't ask.

In summary, during the period 13th of November 1991 to 5th of December 1991 the service received a total of twenty-two calls.

The average length of a wizard operated call was two minutes and approximately three minutes for the human operated calls.

A preliminary analysis of the transcripts produced a variety of interesting statistics on caller behaviour. In particular it was found that, on average, there were signifi-

Wizard Operator:	12 good calls 3 hung up 1 gave up
Human Operator:	6 good calls

Table 1: Summary of the calls.

cantly fewer words spoken by the caller in each turn of the human-wizard condition than in the human-human condition. Also, although the rate of "uhms" and "errs" was about the same in both conditions, callers seemed to be more polite to the machine than to the human operator!

CALLER BEHAVIOUR	Wizard	Human
Average no. of turns/call:	7	18
Average no. of words/call:	31	142
Average no. of words/turn:	4.2	7.9
Turns with "uhms/errs":	12%	14%
Turns with "please/thankyou":	27%	7%

Table 2: Preliminary analysis of the transcripts.

There were two interesting exceptions to the general statistics and the corresponding data was excluded from the analysis.

In the first exception, the behaviour of one caller to the wizard was very different to the other human-wizard interactions but very much in line with the human-human dialogues. A check of the corresponding transcripts revealed that the following utterance was spoken by the caller at the start of the call - "Oh, sorry, it sounded

like a machine that was talking" - after which the caller appeared to continue with the assumption that he was talking to a human operator (despite the peculiar voice)!

In the second exception, another caller, this time to the human operator, exhibited behaviour which was more in line with the human-wizard dialogues. In this case it transpired that the operator had forgotten to switch off the voice disguise unit at the very beginning of the call. Thus this caller seemed to believe he was talking to a machine even after the operator's voice suddenly returned to normal!

7. CONCLUSIONS

This paper has described SRU's first experiences collecting unscripted speech data using the Wizard of Oz technique to provide a *genuine* telephone-based route planning service. Although only a limited quantity of data has been collected so far, several valuable insights into the nature of future speech-based human-machine interaction have been obtained. In particular, various practical issues have been highlighted such as the need to handle significant background noises, spoken asides by the user and interruptions. Also, genuine spoken human-machine interaction appears to be shorter and more (short-term) goal directed than corresponding human-human dialogue.

The service has currently been suspended temporarily with a view to launching a much larger exercise sometime in the early spring of 1992.

References

1. Fraser, N. M. and Gilbert, G. N. "Simulating speech systems", Computer Speech and Language, Vol.5, No.1, pp 81-99, January 1991.

2. Moore, R. K., Tomlinson, M. J. and Morris, A. "Whither the wizard?", Proc. ESCA workshop on the Structure of Multimodal Dialogue, Maratea, Italy, 16-20 September 1991.

3. Polifroni, J., Seneff, S. and Zue, V. "Collection of spontaneous speech for the ATIS domain and comparative analyses of data collected at MIT and TI", Proc. DARPA Speech and Natural Language Workshop, pp 360-365, Pacific-Grove, CA, 19-22 February 1991.

4. Spitz, J. "Collection and analysis of data from real users: implications for speech recognition / understanding systems", Proc. DARPA Speech and Natural Language Workshop, pp 164-169, Pacific-Grove, CA, 19-22 February 1991.

5. Taylor, M. M. *personal communication*, September 1991.

6. Zue, V., Daly, N., Glass, J., Leung, H., Phillips, M., Polfroni, J., Seneff, S. and Soclof, M. "The collection and preliminary analysis of a spontaneous speech database", Proc. DARPA Speech and Natural Language Workshop, pp 126-134, Harwichport, MA, 15-18 October 1989.