RailTel: Railway Telephone Services*

R. Billi and L.F. Lamel

June 29, 1997

1 Introduction

The aim of the LE-MLAP project RailTel (Railway Telephone Information Service) was to evaluate the technical adequacy of available speech technology for interactive telephone services, taking into account the requirements of the service providers and the users. In particular the consortium evaluated the specifications and potential for vocal access to rail travel information. The project, coordinated by CSELT (Italy), included two user organizations: British Rail/ British Systems (U.K.) and FS (Italy). The French railways (SNCF) was a partner in the MAIS project[1] but contributed to the RailTel specification of user and application requirements, and the evaluation methodology. Three research centers, CCIR (U.K.), CSELT, LIMSI-CNRS (France) provided the spoken language technologies, and service providers Saritel and LTV (Italy) contributed expertise concerning telephone services and management of railway timetable databases.

A market study was carried out in coordination with the MAIS project to estimate the current usage of operator-based telephone information services, as well as the expected demand and potential for automated services. Based on a study of 130 information offices in 6 countries, over 100 million calls are handled per year, and there are at least an additional 10 million potential calls which go unanswered. 58% of the callers wait less than 30 seconds and 12% wait more than 1 minute. 91% of the callers ask only for information (97 million calls), with 9% also making a reservation. It is estimated that over 90% of the information calls could be served by an automatic system that could recognize 400 city names (and over 95% with 500 names). Automatic services can thus provide an economic solution to reducing the waiting times and extending the hours of service availability, particularly in light of the fact that in most countries it is the information provider, not the client, that pays for the call. It was found that there is quite a large variability in the services offered in the different countries.

The study also characterized the needs and habits of the users. The vast majority of calls are from individuals, for either business or leisure travel. (Most tourist agents access timetable information electronically.) From the user's perspective the service must be easy to use even for naive users, and any normal speaker should be able to interact with the service. The service should be available via a unique country-wide telephone number, with an automatic transfer to a human operator service in case of difficulty.

*This work was partially financed by the LE MLAP project 63-022 RailTel.

[1]The RailTel project worked in close cooperation with the LE-MLAP project MAIS, that addressed the same task for a different geographical area.

A field trial was carried out in each country to evaluate the performance of the prototype automated service and its acceptability by the end user, as well as to indicate necessary improvements of the system. The field trial methodology was jointly defined by the RAILTEL partners for the 3 prototype systems (Railtel, 1995a,b). The use of a common methodology was necessary in order to be able to compare results of the field trials with different systems in different countries. A total of 100 subjects per language were recruited as volunteers.

Each subject was asked to make a single call to the system and to complete an enclosed questionnaire immediately after interacting with the system. Two types of scenarios were used. In the first type the two cities were connected by a direct train, and the user was supplied with an exact date and time of travel. These scenarios represent relatively simple, but frequent, information requests. In the second type of scenario, the time and date of travel were specified only in general terms, allowing the user more flexibility in formulating the request, as well as a range of interpretations. Travel between the two cities also required changing trains, so as to assess the response generation and synthesis components. For each kind of scenario, several different formulations and a variety of cities, dates and times were used.

Both objective and subjective evaluation measures were defined. The objective performance measures concerned the overall call duration and the number of turns, the dialog success rates, the stage of dialog failure and as well as performance measures for the system components.

The user questionnaire, elaborated by the consortium, contained 20 commonly agreed statements to assess user's subjective impressions and opinions about the system. The polarity of the statements were balanced for negative and positive assessment. In addition to the standard questionnaire, subjects were asked what they considered the good aspects of the system, how it should be improved, and whether they would use such a potential service. Information was also obtained about the subject's travel habits (how often they travel by train, how they obtain their ticket) and their computer experience.

Although the prototypes were at different stages of development, all of them were sufficiently advanced to get initial feedback from end users in terms of functional usability. The presence of moderate speech recognition errors (word error rates up to 25%) did not prevent users from getting the information for their intended scenarios.

The first type of scenario had a higher completion rate, and shorter overall duration than the second type. The longer dialogs for the second scenario types are mainly due to the refinement of travel times, as well as to longer system responses needed to provide information on changing trains.

The usability profiles for the three prototypes are shown in Figure 1. The highest subjective scores were for ease of use (Q1), information provided (Q18), friendliness (Q3); the lowest were for speed (Q19), reliability (Q6), preference for human operator (Q13), need for improvement (Q16). Most of the callers succeeded in obtaining the travel information for their scenarios; the questionnaire responses for these callers were broadly favorable towards the systems. Some calls were unsuccessful for a variety of reasons like recognition failures and reliability of the platform.

The overwhelming positive responses suggest that the basic design of a simple system-directed dialogue is usable if the dialogue is carefully designed. More sophisticated natural language understanding system were also well received. The field trial revealed areas for improvement in the speech recognition and interface design, in the database search and in

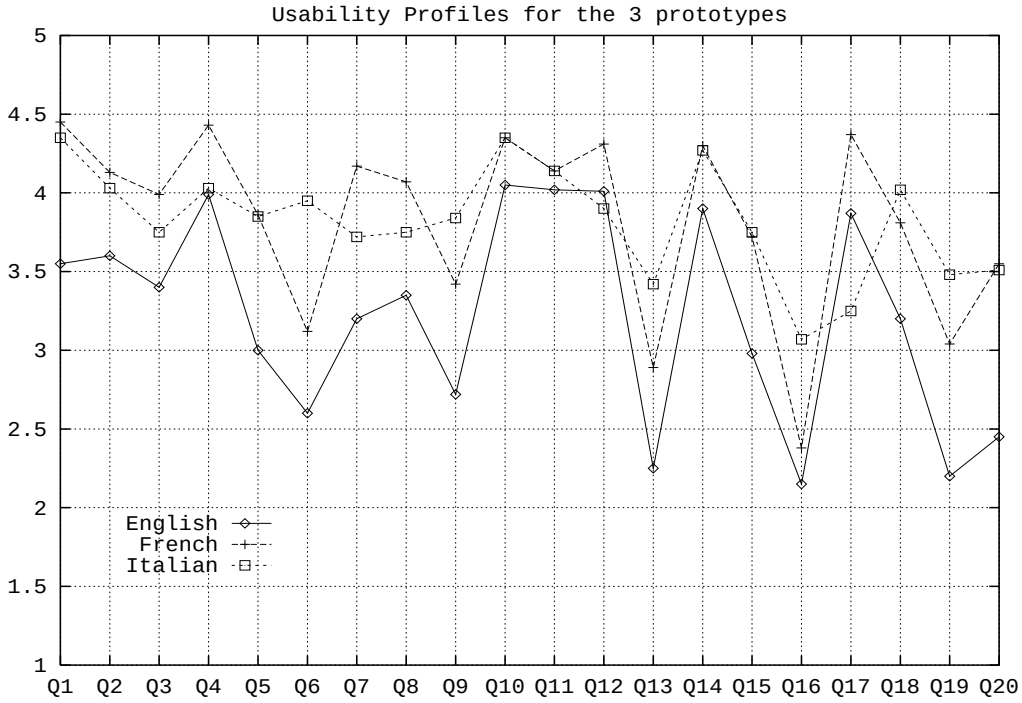

Figure 1: Overall results for usability profiles as a function of scenario type. Q1: ease-of-use, Q2: confusing, Q3: friendliness, Q4: complex, Q5: use again, Q6: reliability, Q7: control, Q8: concentration, Q9: efficiency, Q10: fluster, Q11: too fast, Q12: stress, Q13: prefer human service, Q14: complicated, Q15: enjoyable, Q16: needs improvement, Q17: politeness, Q18: information obtained, Q19: faster than human, Q20: understood

the reliability of the platform on which the prototype systems were implemented. The test bed developed in the project will permit us to evaluate speech technology improvements and to carry out more extensive tests with real users.

The following two papers give a description of the spoken language systems developed for the Italian and French languages, and the corresponding field trials carried out according to the commonly agreed upon protocols. For the Italian language two different dialog strategies are contrasted: a system-directed dialog, and a mixed-initiative dialog.

References

RailTel (1995a), "Definition of the evaluation methodology for the Field Trials," RailTel/Mais *Project deliverable D4*, Saritel, June.

RailTel (1995b), "Results of Field Trials," RailTel *Project deliverable D8*, November.

List of Figures