# Field Trials of a Telephone Service for Rail Travel Information*

*L.F. Lamel, J.L. Gauvain, S.K. Bennacef, L. Devillers, S. Foukia, J.J. Gangolf, S. Rosset*
Spoken Language Processing Group
LIMSI-CNRS
91403 Orsay, FRANCE
lamel@limsi.fr

**Abstract –** This paper reports on the RAILTEL field trial carried out by LIMSI, to assess the technical adequacy of available speech technology for interactive vocal access to static train timetable information. The data collection system used to carry out the field trials, is based on the LIMSI MASK spoken language system and runs on a Unix workstation with a high quality telephone interface. The spoken language system allows a mixed-initiative dialog where the user can provide any information at any point in time. Experienced users are thus able to provide all the information needed for database access in a single sentence, whereas less experienced users tend to provide shorter responses, allowing the system to guide them. The RAILTEL field trial was carried out using a commonly defined upon methodology. 100 naive subjects participated in the field trials, each contributing one call and completing a user questionnaire. 72% of the callers successfully completed their scenarios. The subjective assessment of the service was for the most part favorable, with subjects expressing interest in using such a service.

## I. INTRODUCTION

In this paper we report on the RAILTEL field trials carried out by LIMSI. The LE-MLAP project Railway Telephone Information Service (RAILTEL) aimed to evaluate technical adequacy of available speech technology for interactive telephone services, in particular the potential for vocal access to rail travel information. A particularity of telephone information services is that all interaction with the user including all information returned by the system, must be exchanged vocally, making oral dialog management and response generation very important aspects of the system design and usability.

The RAILTEL data collection spoken language system is largely based on the spoken language system developed for the ESPRIT MASK project[8]. The system runs on a Unix workstation with a high quality telephone interface which can support up to 4 telephone lines. The LIMSI prototype service was developed over the summer of 1995, and demonstrated at the *Eurospeech'95* conference. This system was used to collect telephone data (over 4000 queries) with which new acoustic models and language models were constructed for the speech recognizer. The prototype service was used to carry out a field trail with 100 naive users during the fall of 1995, according to a common field trial protocol designed for the project. Field trials were carried out by out Italian partners and British partners with their prototype systems.

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

## II. RAILTEL DATA COLLECTION SYSTEM

An overview of the spoken language system[13] for information retrieval is shown in Figure 1. The main components are the speech recognizer, the natural language component which includes a semantic analyzer and a dialog manager, and an information retrieval component that includes database access and response generation. While our goal is to develop underlying technology that is speaker, task and language independent, any spoken language system will necessarily have some dependence of the chosen task and on the languages known to the system[12]. The spoken query is decoded by a speaker independent, continuous speech recognizer[9], whose output is then passed to the natural language component. In our current implementation the output of the speech recognizer is the best word sequence, however, the recognizer is also able to provide a word lattice. The semantic analyzer carries out a caseframe analysis to determine the meaning of the query, and builds an appropriate semantic frame representation[4]. The dialog history and default values generated from the task knowledge are used to complete missing information in the semantic frame. Should additional information be required for database access, the dialog manager prompts the user to fill in missing information, and then generates a database query. The system accesses a copy of the static train information (database RIHO) via a network connection. The returned information is converted to a natural language response by the response generator, which is synthesized by speech concatenation and played to the user.

### *A. Speech Recognition*

The speech recognizer is a medium vocabulary, real-time, speaker-independent, continuous speech recognizer. Speaker independence is achieved by using acoustic models which have been trained on speech data from a large number of representative speakers, covering a wide variety of accents and voice qualities. The recognizer uses continuous density HMM with Gaussian mixture for acoustic modeling and *n-gram* backoff language models[10]. Context-dependent phone models are used to account for allophonic variation observed in different contextual environments. The *n*-gram statistics are estimated on the transcriptions of spoken queries. Since the amount of language model training data is small, some grammatical classes (such as cities, days, months, etc) are used to provide more robust estimates of the *n*-gram probabilities. The current RAILTEL recognition vocabulary contains about 1500 words, including the 600 station/city names specified by the SNCF. The recognition vocabulary used in the field trials contained 800 words, including 58 station names.

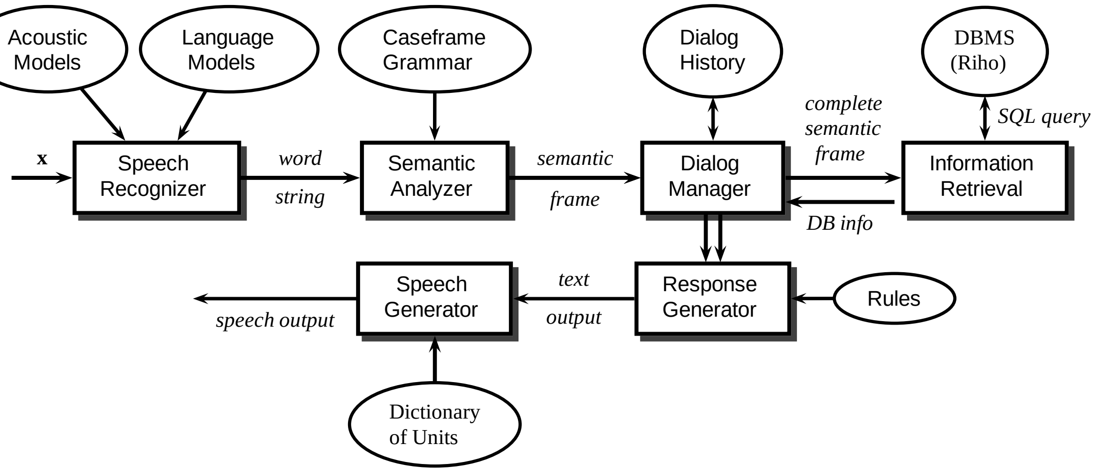


**Figure 1:** Overview of the RAILTEL data collection system for spoken language information retrieval. **x** is the input speech signal.

## B. *Natural Language Understanding*

The semantic analyzer carries out a caseframe analysis[7, 6] to determine the meaning of the query[4], and builds an appropriate semantic frame representation. The resulting semantic frame, which contains a set of slots instanciated by the meaningful words of the utterance, we refer to as "literal understanding"[3]. In the caseframe analysis, keywords are used to select an appropriate case structure for the query without attempting to carry out a complete syntactic analysis. The major work in developing the understanding component is defining the concepts that are meaningful for the task and the appropriate keywords. The concepts for the RAILTEL task are **train-time**, **fare**, **change**, **type**, **reserve**, **service** and **reduction**. While in the RAILTEL field trials only a subset of these concepts were directly used (**train-time** and **change**), subjects solved additional more complicated scenarios to test the system capabilities.

## C. *Dialog Management*

The dialog manager ensures the smooth interface between the user and the computer. The dialog consists of three phases[5, 3]: main information exchanges preceded and closed by formalities; and is structured into a hierarchy of sub-dialogs with different types of responses associated with the different sub-dialogs. The dialog manager maintains the dialog history which is used to complete missing information in the semantic frame. The dialog context may be used to provide default values for required slots. We refer to the reinterpretation of the query in the context of the ongoing dialog as "contextual understanding"[3]. A mixed-initiative dialog strategy is used where the user is free to ask any question, at any time. However, in order to aid the user, the system prompts the user for any missing information needed for database access. Experienced users are able to provide all the information needed for database access in a single sentence, whereas less experienced users tend to provide shorter responses, allowing the system to guide them.

The completed semantic frame is used to generate an SQL-like request to the database management system, RIHO. Interpretative and history management rules are applied prior to generation of the DBMS request, and post-processing rules are used to interpret the returned information prior to presentation to the user. For example, in order to provide a more cooperative dialogue and response, the system relaxes constraints on the departure time when no train corresponds to the user's request, allowing the system to return the closest train after or before the specified time.

Since there is no visual support in the telephone communication, response generation plays a very important role in the overall system. The generation of responses is complex because if too much information is given, it may be difficult for the user to extract the important part. If not enough information is returned, the interaction will take longer, as the user will need to ask for more detailed or additional information. The system responses depend on the dialog context and on the information returned from the database management system. Careful attention has been paid to construction of sentences that contain the appropriate information and the generation of natural-sounding utterances[3]. Messages are synthesized by concatenation of speech units stored in the form of a dictionary. There are about variable-sized 2000 units for the RAILTEL application.

# III. FIELD TRIAL METHODOLOGY

The field trial methodology was jointly defined by the RAILTEL partners for the 3 prototype systems[1]. The LIMSI RAILTEL information system was accessible 24 hours a day via a toll-free number. A single telephone line was used. For legal reasons, a recorded message is presented at the start of each call, informing the caller that their voice will be recorded for the purposes of research and development, and that if they do not agree to be recorded, they should hangup.

A total of 100 subjects were recruited for the field trial from 3 sources: 77 of the subjects were recruited by LIMSI (they responded to a newspaper announcement and were paid for their participation), the remaining subjects were employees or family members of the SNCF (14 callers) or the Vecsys company (9 callers). Each subject was asked to make a single call (a single scenario of type **A** or type **B** as shown in Figure 2), and to complete the enclosed questionnaire immediately after interacting with the system. Scenarios of type **A** supply the user with an exact date and time of travel, and represent relatively simple, but frequent, information requests. In scenario **A**

**Scenario A**
You want to find out the departure time of a train from *[city_A]* to *[city_B]*, on [*date*] at [*time*].

(You want to take a direct train from Paris to Bordeaux on March 14th leaving at 9 am.)

*Note that* city_A *and* city_B *must be connected by a direct train, and that the time and date of travel are specified.*

**Scenario B**
Find the arrival time of an *[time-period]* train from *[city_A]* to *[city_B]*, next *[relative-date]*.

(You would like to know the arrival time of an evening train from Lyon to Grenoble next Wednesday.)

*Note that traveling from* city_A *to* city_B *must be require a change of trains, and that the time and date of travel are specified only in general terms.*

**Figure 2:** Commonly defined scenarios used in the field trials.

*city_A* and *city_B* must be connected by a direct train. The scenarios of type **B** allow more flexibility on the part of the user, as well as a range of interpretations since the time and date of travel are specified only in general terms. The constraint that the trip require changing trains is to assess the response generation and synthesis components. For each kind of scenario, at least six different formulations were used. Combined with the different town names, dates and train times, we generate a large set of different scenarios.[1]

Each subject completed a questionnaire to gather their immediate impression of the prototype system. The questionnaire, elaborated in coordination with the other partners contains 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, we asked subjects 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.

## IV. Field Trial Results

The field trial results are based on the first 50 calls of each type for which a completed questionnaire was returned. Table 1 provides general information about the callers. Although no specific selection was made to balance gender or age, there are roughly 50% callers of each sex. There are more male callers for scenario **A** and more female for scenario **B**. 36% of the callers are younger than 25, and 7% are older than 50. The recruitment origin of the callers may also reflect their experience. For example, those recruited by LIMSI are not predisposed to have any experience with vocal servers, computers nor any particular travel habits. The 14 subjects recruited by the SNCF

| | *Sex* | | *Age* | | |
|---|---|---|---|---|---|
| *Scenario* | *male* | *female* | $< 25$ | $25 - 50$ | $> 50$ |
| **A** | 30 | 20 | 15 | 31 | 4 |
| **B** | 22 | 28 | 21 | 26 | 3 |
| **A+B** | 52 | 48 | 36 | 57 | 7 |

**Table 1:** Field trial sample overview: gender and age of subjects.

can be expected to have a good knowledge of the rail system, and to be frequent travellers. Those recruited by VECSYS may be expected to be somewhat familiar with computers and may have had experience with voice technology.

### A. Global Evaluation

The data registered for each call concern the call duration and the number of turns, as shown in Table 2. The average dialog duration is 193 secs for type **A** and 245 secs for type **B** scenarios. The longer duration is correlated with the larger number of turns for the type **B** scenarios (5) than for type **A** (4). This is due primarily to the refinement of the time for scenario **B**, which is specificied in general terms. For the 50 type **A** scenarios 76% of calls were successfully completed, compared to 68% success for scenarios of type **B**.

| *Scenario* | *Turns per call* | *Duration* | *Success Rate* |
|---|---|---|---|
| **A** | 3 | 193 secs | 76% |
| **B** | 5 | 245 secs | 68% |
| **A+B** | 4 | 219 secs | 72% |

**Table 2:** Objective measure for field trial data.

For the purposes of comparison with the Italian and UK demonstrators we evaluated the performance as a function of the numbers of calls reaching specified stages in the dialog as shown in Table 3. A call is counted as failing at a particular stage if the final response of the system does not correspond to the original request for this stage.[2] More than one stage could be a failure for a given dialog. The order of stages is irrelevant in our system as information can be specified in any order.

| *Scenario* | *Start* | *DepCity* | *ArrCity* | *Date* | *Time* | *Play* |
|---|---|---|---|---|---|---|
| **A** | 0% | 6% | 6% | 0% | 8% | 4% |
| **B** | 2% | 10% | 6% | 8% | 6% | 2% |
| **A+B** | 1% | 8% | 6% | 4% | 7% | 3% |

**Table 3:** Number of calls failing at specified stages in the transaction. The stages are defined as: **Start:** Travel demand origin entered; **DepCity:** Departure-City recognised and understood; **ArrCity:** Arrival-City recognised and understood; **Date:** Date of travel recognised and understood; **Time:** Time of travel recognised and understood; **Play:** Response of response generator.

Overall, in 34.8% of the queries, there was a recognition error. These errors do not necessarily result in a failed scenario (28%). For the unsuccessful scenarios, 80% fail due to recognition and understanding errors, 14% can be contributed to the dialog management, and the remaining 6% are a result of information retrieval errors. Databases

[1]The subjects recruited by LIMSI completed 5 calls to the system, the first call serving for the field trial. Three extra scenarios were designed for data collection purposes changing the presentation style, and asking callers to find out information about concepts not yet handled by the system. This enabled us to collect data for a wider variety of situations, and to see how users reacted when the system was unable to provide them the information they wanted, such as for example, when a station or city-name was not known to the system This data will help us to develop ways to detect such situations.

[2]The notion of stage in our system is not the same as for the Italian and UK demonstrators. Since our system uses a mixed-initiative dialog, the machine does not control the order in which information is supplied by the user. Only when additional information is needed for database access does the system lead the dialog, and even in these cases, the user may choose to provide other information than what is asked for, and this information is taken into account.

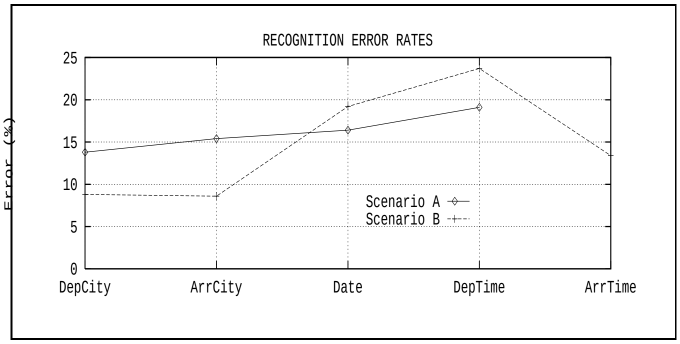


**Figure 3:** Recognition error rate as a function of slot type.

access was not specifically evaluated in the field trials, this step was assumed to be error free.

A multilevel error analysis has been carried out of the field trial system distinguishing errors due to recognition, understanding and to dialog. Each level is evaluated by differentiating errors caused at the current level from errors propagating from the lower levels. Thus, to evaluate the understanding level we separate out errors due to recognition errors from those attributable to the understanding component. Similary, the dialog is evaluated by differenciating between errors due to the recognition and understanding levels and those arising from the dialog level.

## *B. Speech Recognition Performance*

The speech recognition component was evaluated on an independent set of test sentences, and has a word error of about 18%. However, this number is can be misleading as the word accuracy measures all differences between the exact orthographic of the query and the recognizer output. Many recognition errors (such as masculine/feminine forms, or plurals) are not important for understanding.

Therefore, we evaluated the recognition performance for the slots relevant for the understanding component. There were a total 1284 input attempts in the 100 calls. 16 of the inputs were rejected (1.2%). Of these 6 were empty and 2 where the hangup beep. For the remaining queries the percentage of slots not recognised are shown in Figure 3 for the two scenario types. In each case, the number of erroneous slot instantiations are divided by the total number of instantiated slots of that type after literal understanding. The number of slot instantiations for the different slot types are given in Table 5. Due to the scenario definition, there were no **ArrTime** instantiations for the type **A** scenarios.

The slot recognition error is on the order of 10-15% for **DepCity** and **ArrCity** and include errors due to misrecognition of the actual city name and also errors on premarkers signaling "to" or "from". For example, during the field trial we observed for the first time the formulation "à destination de Paris" which instantiated the slot **Departure-City:** Paris instead of the slot **Arrival-City** (see Figure 4). For dates and times the main errors are due to the insertion of extra digits, such as "12:30" (douze heure trente) being recognized as "12:37" (douze heure trente sept). The type **A** scenarios had more recognition errors on cities, while type **B** had more errors on dates and times.

## *C. Spoken Language Understanding Performance*

To evaluate the understanding performance it is necessary to differenciate between errors due to recognition errors and errors due to

| *Scenario Type* | *Recognition* | *Understanding* |
|---|---|---|
| **A** | 23.2% | 10.7% |
| **B** | 20.0% | 6.0% |
| **A+B** | 21.6% | 8.4% |

**Table 4:** Average semantic frame slot understanding error rate.

**U:** [BB] je souhaiterais un train au départ d' Orléans et *à* destination de Paris le 17 mars vers 7 heures (0l8002)
**R:** je souhaiterais un train au départ d' Orléans et destination de Paris le dix sept mars vers sept heures (0l8002)
<train-heure>
{ from: orléans. (1)
from-stand: paris. (1)
heure-depart-relative: vers. (1)
heure-depart-heure: 7. (1)
jour-depart-jour: 17. (1)
jour-depart-mois: mars. (1)
depart: part. (1)
}

**Figure 4:** Recognition error that does not cause an understanding error. **U** is the orthographic transcription of the query, **R** is the recognized word string.

understanding errors. Table 4 shows the recognition and understanding query error rates for scenarios **A** and **B**. These error rates are calculated by averaging the instantiated slot error rates for all queries. For each semantic frame, all slots which are incorrectly instanciated are marked with the error source, recognition or understanding. It is then straightforward to compute the incorrect slot instanciation rate (due to recognition or understanding) for the semantic frame by simply dividing the number slots with error by the total number of instantiated slots.

The query understanding error rate is seen to be about 40% that of the recognition error rate. This is because not all recognition errors lead to an error in understanding. In Figure 4 an example is given of such an error, where the word *à* was deleted. A recognition error on a city name systematically results in an understanding error. These errors are usually corrected in the ensuing dialog and do not cause the dialog to fail, unless for example when the desired city name is outside of the recognition vocabulary.

Table 5 shows the percentage of slots not understood for the two types of scenarios. The per slot understanding error is very low for the departure and arrival cities as almost all of the errors are caused by recognition errors. The understanding errors for arrival and departure times are on the order of 20%. Not all understanding errors are important for dialog success, for example, interpreting time period as "around 10pm" instead of "after 10pm" may not affect the information

| | *Slot Type* | | | | |
|---|---|---|---|---|---|
| *Type* | *DepCity* | *ArrCity* | *Date* | *DepTime* | *ArrTime* |
| **A** *#slots* | 72 | 69 | 202 | 264 | - |
| und | 0.0% | 0.0% | 8.8% | 17.2% | - |
| **B** *#slots* | 98 | 96 | 144 | 191 | 142 |
| (und) | 0.2% | 1.1% | 0.0% | 24.4% | 17.3% |

**Table 5:** Total number of slot instantiations for each slot type after literal understanding and the understanding error rates for each type of slot. Total of 607 slots for type **A** and 671 slots for type **B** queries.

| *Scenario Type* | *# Dialogs* | *Correct* | *Cause of error* | | |
|---|---|---|---|---|---|
| | | | *Rec/Und* | *Dialog* | *DB* |
| **A** | 50 | 58.5% | 34.3% | 0.9% | 6.3% |
| **B** | 50 | 60.5% | 29.2% | 10.3% | - |

**Table 6:** Source of dialog error per system response.

**U:** I'd like to know the next train (0zl005)
`[train already given:  12:33]`
<train-time>
    order-stand: next. (1)
    }
**S:** There are more than 10 trains from Arras to Lille-Flandres Thursday 28/09 corresponding to your request. The first leaves at 13:06 and the last at 21:54. Please give a more specific departure time. (D-0)

**Figure 5:** Example of a dialog error.

obtained from the database, and therefore has no affect on the dialog. It has been our observation that such minor understanding errors pass unobserved by the user, where as more important understanding errors will lead to longer dialogs, as the user tries to correct the error.

To evaluate the dialog performance it is necessary to distinguish between errors due to understanding component and errors due to dialog manager. The Table 6 shows the dialog error rate for both scenario types. Dialog errors due to recognition and understanding errors are more important for type **A** scenarios than for type **B** scenarios. Dialog errors appearing in the type **B** scenarios were linked to the understanding of the arrival time and date, when it was necessary to depart the previous evening to arrive at the specified time. These were not correctly handled in the system and implied dialog failure. This error was actually due to a problem in post-processing the retrieved information, but we have considered it to be a dialog error as the system response was not correct in the context of the user's query. The field trials turned up a database connection problem, that was corrected early on (column DB).

An example of a dialog error is illustrated in Figure 5. This occurred when the user asked for "the next train" (the train already returned by the system was at 12:33). The query was both correctly recognized and understood, but an incorrect response was given to the user. The system returned all trains after 12h33 instead of giving just the next one at 13:06, and asked the the user to specify a more specific departure time.

## *D. User Evaluation*

The responses to 100 questionnaires were used to generate "usability profiles". The overall user assessment on a scale of 5 is shown in Figure 6 individually for the two scenarios types and combined. Both scenarios types are rated at the same level by the subjects. Although not shown in the figures, there is a slight tendancy of younger subjects to asses the system more favorably than the older subjects ($< 25 : 3.8$; $25 <> 50 : 3.7$; $> 50 : 3.5$). This is likely to be correlated with a larger familiarity of younger subjects with computers and automated services. Female subjects tended to assess the system more favorably than the male subjects (see Figure 7), however, they also expressed less motivation to use such a system (Q13). A priori we expected that there may be a difference in user assessment due to the recruitment source. As shown in Figure 8 there does not seem to be much of a difference. The 9 VECSYS subjects have slightly more extreme ratings, but the differences are not significant.

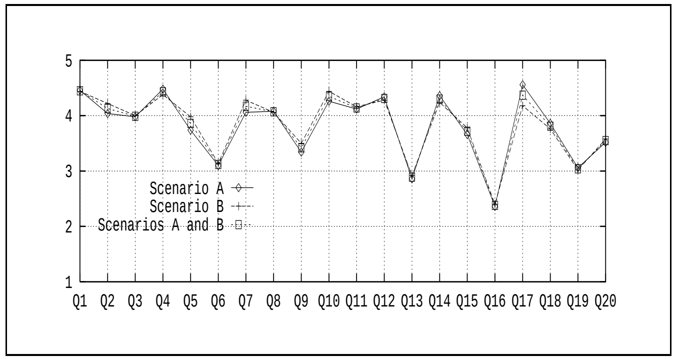


**Figure 6:** Overall results for usability profiles as a function of scenario type.Q1: ease-of-use, Q2: confusing, Q3: friendliness, Q4: complex, Q5: use again, Q6: reliability, Q7: control, Q8: concentration, Q9: efficiency, Q10: fluster, Q11: too fast, Q12: stress, Q13: prefer human service, Q14: complicated, Q15: enjoyable, Q16: needs improvement, Q17: polite , Q18: information obtained, Q19: faster than human, Q20: understood

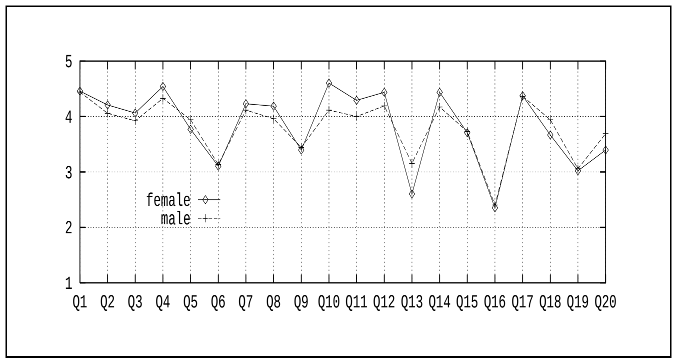


**Figure 7:** Usability profiles for male and female callers.

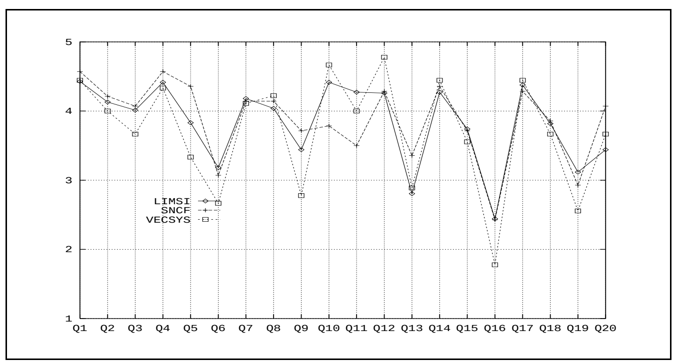


**Figure 8:** Usability profiles for callers by recruitment sources.

We grouped the questions into the following 5 categories: Attitude (A: 5, 12, 13, 15), ease of use (EU: 1, 2, 4, 8, 14), efficiency (E: 9, 16, 19), reliability (R:6, 18, 20) and user-friendliness (UF: 3, 7, 10, 11, 17). Figure 9 shows the overall results for these 5 categories. While there is a tendancy of subjects to assess the system favorably (EU and UF), they don't find it particularly efficient (E), and some subjects

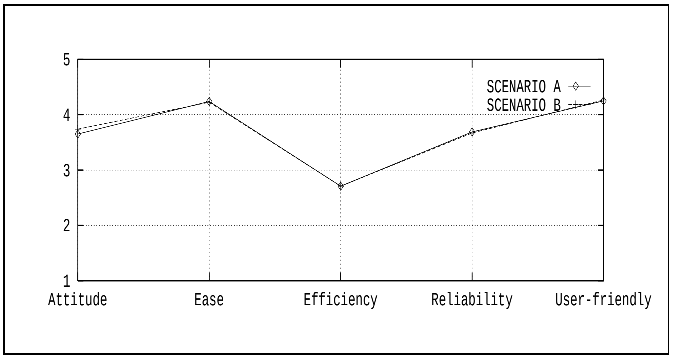


**Figure 9:** Overall results for the 5 categories.

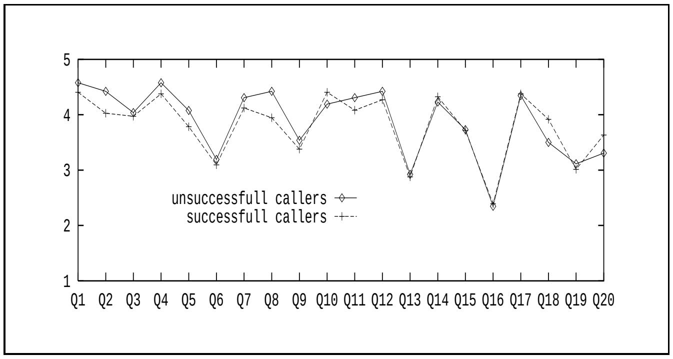


**Figure 10:** Results for successful and unsuccessful callers

doubted the reliability of the system (R). This is likely to be linked to the responses to question Q16, that the system needs to be improved.

In Figure 10 the responses of subjects are differentiated into groups of successful and unsuccessful callers. We observe that these two groups rate the system at the same level.

It is possible to relate these results with the two last questions we added to the questionnaire - subjects were asked what are the good aspects of the system and how it should be improved. Concerning the negative points, subjects expressed mainly concern about the reliability of the system and the information returned. Concerning the positive points, subjects commented on the user-friendliness and the speed of the system, several judging it to be faster than to use the human service.

## V. DISCUSSION AND SUMMARY

We have described our data collection system for access to train travel information over the telephone. A preliminary system, based on our MASK system[8], was brought up very quickly so as be able to carry out the RAILTEL field trial. The data collected with 100 naive subjects are now being used to improve the system capabilities. The performance snapshot resulting from the field trial had 72% of the callers successfully completing their scenarios. The failures were due mainly to recognition and understanding errors (80%), with 14% due to dialog management, and the remaining 6% resulting from information retrieval errors.

Subjects rated the system at the same level for both types of scenarios (Figure 6) even though the type **A** scenarios are easier, and the dialogue failure rates are quite a bit different. The reason could be that we have considered a dialog to be a failure even if final response of the system was only slightly different from what was specified in the scenario. This judgement was used even if the subject did not exactly respect the scenario. Therefore subjects may have been happy with a response that we considered erroneous.

A correlation between objective measurements and the subjective assessment (via the questionnaire) was observed with respect to the age of subjects. Older subjects ($> 50$) expressed less satisfaction with the system, and had higher dialog error rates. This can be partly contributed to the lack of experience of older users with computers and automated services, and also that our training corpus does not include much data from older speakers.

The subject assessment of the service was largely favorable, although there was a clear expression of the need for improvement. Most subjects expressed a potential interest in using such a service. The field trial demonstrates that such services should be easily accepted by the general public and that further developments to put together a real service are worth pursuing.


## ACKNOWLEDGMENT

We would like to thank the SNCF for providing the information database RIHO for use in the RAILTEL project.


## REFERENCES


[1] "Definition of the evaluation methodology for the Field Trials," RAILTEL/MAIS *Project deliverable D4*, Saritel, June 1995.

[2] "Results of Field Trials," RAILTEL *Project deliverable D8*, Nov. 1995.

[3] S.K. Bennacef, L. Devillers, S. Rosset, L.F. Lamel, "Dialog in the RAILTEL Telephone-Based System," to appear in *ICSLP'96*, Oct. 1996.

[4] S.K. Bennacef, H. Bonneau-Maynard, J.L. Gauvain, L.F. Lamel, W. Minker, " A Spoken Language System For Information Retrieval," *ICSLP'94*, Yokohama, Japan, Oct. 1994.

[5] S.K. Bennacef, F. Neel, H. Bonneau-Maynard, "An Oral Dialogue Model based on Speech Acts Categorization," *ESCA Workshop on Spoken Dialog Systems*, Vigsø, Denmark, Spring 1995.

[6] B. Bruce, "Case Systems for Natural Language," *Artificial Intelligence*, **6**, 1975.

[7] Ch.J. Fillmore, "The case for case," in *Universals in Linguistic Theory*, Emmon Bach & Robert T. Harms (eds.), Holt, Rinehart and Winston, Inc., 1968.

[8] J.L. Gauvain, S.K. Bennacef, L. Devillers, L.F. Lamel, S. Rosset, "The Spoken Language Component of the Mask Kiosk," *Proc. Human Comfort & Security Workshop*, Brussels, Oct. 26, 1995.

[9] J.L. Gauvain, L.F. Lamel, G. Adda, M. Adda-Decker "Speaker-Independent Continuous Speech Dictation," *Speech Communication*, **15**, pp. 21-37, Sept. 1994.

[10] S.M. Katz, "Estimation of Probabilities from Sparse Data for the Language Model Component of a Speech Recognizer," *IEEE Trans. ASSP*, **35**(3), 1987.

[11] L.F. Lamel, J.L. Gauvain, B. Prouts, C. Bouhier, R. Boesch, "Generation and Synthesis of Broadcast Messages," *Proc. ESCA-NATO Workshop on Applications of Speech Technology*, Lautrach, Germany, Sept. 1993.

[12] L.F. Lamel, S.K. Bennacef, H. Bonneau-Maynard, S. Rosset, J.L. Gauvain, "Recent Developments in Spoken Language Sytems for Information Retrieval," *Proc. ESCA Workshop on Spoken Dialog Systems*, Vigsø, Denmark, Spring 1995.

[13] L.F. Lamel, S. Rosset, S.K. Bennacef, H. Bonneau-Maynard, L. Devillers, J.L. Gauvain, "Development of Spoken Language Corpora for Travel Information", *Eurospeech'95*, Madrid, Spain, Sept. 1995.