ISCA Archive
http://www.isca-speech.org/archive

EUROSPEECH '97
5th European Conference on Speech
Communication and Technology
Rhodes, Greece, September 22-25, 1997

ACCeSS - Automated Call Center Through Speech Understanding System

Ute Ehrlich, Gerhard Hanrieder♦, Ludwig Hitzenberger♦♦, Paul Heisterkamp,
Klaus Mecklenburg, Peter Regel-Brietzmann,

Daimler-Benz AG, Institute for Information Technology
Wilhelm-Runge-Str. 11, D-89081 Ulm, Germany
regel@dbag.ulm.daimlerbenz.com
♦ Daimler-Benz Aerospace, Wörthstr. 85, D-89077 Ulm, Germany
♦♦ University of Regensburg, FG Information Science, D-93040 Regensburg, Germany

ABSTRACT

This paper will describe the results of a high sophisticated speech application project. ACCeSS[1] is an EU project with Greek and German partners. Our Greek partners are Knowledge S.A. and the University of Patras. In this paper we report about the German application of the project. The project addresses a first step for automation of call centers for personal intensive applications in insurances. New forms of insurance operation are more and more using the telephone or direct mailing for the contact between an insurance and its customers. This makes the business more efficient and more direct than the classical operation with agents. Routine contractual details can easily be handled in such a way and all other cases of insurance actions can be realised with such communication media, too. A direct insurance company running a large call center is involved in the project and pays attention to the functionality of the system. The user needs are analysed using Wizard-of-Oz experiments. The system structure, a sketch of the algorithms and modules, and first results of evaluation will be given.
Keywords: speech dialogue, Call Center, Wizard-of-Oz, dialogue strategy, semantics, ACCeSS

1. Functionality

The goal of the ACCeSS project is to provide two application partners, a Greek insurance with a net of agents and a German direct insurance doing business over telecommunication lines only, with a solution that helps them to largely automatise their respective call centres. We concentrate here on the application for the direct insurance, but most of what we present applies mutatis mutandis also or the Greek partner.
The direct insurance‘s business area is mostly car insurance. Customers call in or send a fax to get a price offer. These calls average at about 1.500/day, but have extreme peaks during advertising campaigns (two to three p. a.) and immediately after TV commercials. ACCeSS will provide the technology to handle this immense number of calls through speech dialogue.
Direct insurances use exclusively the telephone or direct mailing for the contact to its customers. Up to the moment all the call center work is done by human operators. Customers should get an efficient and fast access at every time to the services of the insurance and can even get some relevant information over the weekend and during the night. One important point is to equalise the loads of the call center. In rush hours (e.g. at lunch time) the heaps of incoming calls can not be managed by the operators and some calls are lost. Our paper describes the (partly) automation of customer acquisition. We will only describe the call center automation through a speech dialogue system, the automatic fax analysis will not fit the conference topics.

2. Wizard-of-Oz Experiments

For the design of man-machine communication systems the normal human-to-human communication cannot be used as a prototype, because man-machine communication has its own peculiarities (cf. *Krause/Hitzenberger 1992* & *Hitzenberger/Kritzenberger 1989*). To learn about the behaviour of automated call centers we performed several wizard-of-oz experiments with a simulated one. The aim of these experiments was to get indications about the use of language, the structure of dialogues, the acceptability of unavoidable restrictions, the ability of users to deal with spelling alphabets, the operation of feedback mechanisms (echo) and the general acceptance of callers towards talking to an automated system instead with a human operator.

EXPERIMENTAL SET-UP: In the wizard-of-oz situation it is important, that the system behaviour be as close to reality as possible. So we built an experimental setting, which enables the wizard to answer, with the aid of pre-recorded sound files, nearly in real-time as well as to react to unforeseen user input with direct speech. All speech was sent through the same technical environment including a distortion device to produce an even degree of distortion, something which users expect in a computerised speech system. The test persons called the system with the aid of a normal telephone device. The layout for the logical system behaviour was acquired through a thorough analysis of real-life acquisition dialogues in an existing call center.
In a semi-structured interview before and after the experiments the test persons were asked about their attitudes and feelings toward the use of such automated systems.

RESULTS: Generally speaking, the attitude of the test persons towards the system was relatively friendly but critical. The usage of the language and the structure of the dialogues was significantly simpler than in the human dialogues. The average

[1] The ACCeSS project is supported by the Language Engeneering sector of the Telematics Application Program of the European Union. Reference Number LE1-1802.

length of the utterances was mostly between two and three words and the number of clarification dialogues was significantly lower than with the human operators. The test persons had no problems in accepting the initiative of the system in guiding the course of the dialogue, but we encountered problems with the confirmation by system echo. Some of the test persons did not realise that they had the possibility to correct their utterances and thus accepted wrong entries.

In general, the system restrictions were handled properly by the test persons. One of the most important of these was the requirement to use a spelling alphabet. This spelling mode was correctly used most of the time, but in the case of incorrect utterances, the readiness to repeat a wrong word was low. The statistical analysis showed, that there is no significant influence (95% level) of the various echo modes on any of the analysed parameters. There is an influence (90% level) of the echo on the length of utterances, and an influence of different system co-operativity on the handling of the system restrictions.

The acceptance of the system is indeterminate. About half of the test persons complained about the impersonalness of the system, the lack of both clarification possibilities and control. They also criticised the lengthiness and the schematic course of the dialogues. On the other hand most of the test persons did not feel under pressure by giving personal information, although some of them refused explicitly to give this type of data.

3. System Overview

Our philosophy of robust speech dialogue systems is described e.g. in *Brietzmann 1994*. Although most users of speech dialogue systems are co-operative, they are not always experienced in conversing with an automatic dialogue system. Thus in real world applications we have to deal with effects like open vocabulary, sentence breaks and ungrammatical sentences. In addition, a recognition system should be able to deal with different speakers over telephone lines. Therefore, word recognition and linguistic analysis must be very closely co-ordinated and must make use of all available restrictions from higher level modules.

The dialogue system consists of the following components: speech analysis including linguistic analysis, contextual interpretation, dialogue control, the task interpretation, answer generation, and speech synthesis. We have a flexible hierarchical structure that achieves the best overall results for a given utterance. The dialogue control is able to control the speech analysis in several ways: determination of statistical language models, restrictions to the vocabulary, switch to special modes like spelling mode.

LINGUISTIC ANALYSIS (cf. *Ehrlich & Hanrieder 1996*): Linguistic analysis fulfils two very important tasks within the system: On the one hand, it provides top down restrictions for the word recognition stage and reduces input ambiguity on the word level. On the other hand, it discovers the most plausible linguistic interpretation of the input for further dialogue and task-oriented processing. Considering the huge search space needed for robust continuous word recognition, the search in the word graph hypotheses must be performed within one integrated process.

The module is implemented as a constrained based, island driven chart parser using the syntactic and semantic constraints encoded in the underlying unification-based Categorial Grammar (UCG). Robustness is guaranteed by any-time behaviour of the parser: if no syntactically and semantically correct utterance spanning the whole speech signal could be found due to a given time limit for the parsing process, recognition errors, or spontaneous speech effects several partial interpretations are provided for further analysis.

Both predictions by the DIALOGUE CONTROL and the knowledge of the surface structure of the previous system utterance are taken into account in order to reduce the search space.

CONTEXTUAL INTERPRETATION (cf. *Heisterkamp 1993*): The surface-oriented semantic interpretations found by the LINGUISTIC ANALYSIS module are translated into task-oriented representations suitable for the application database, e.g. in a flight information system the time information 'noon' in the caller's utterances 'I have to be at Brussels at noon' and 'The noon flight to Brussels' is mapped to the arrival time and the departure time, respectively. The resolution of deictic (e.g. 'today') and anaphoric expressions (e.g. 'it', 'the previous connection') is also a task of the CONTEXTUAL INTERPRETATION. Caller corrections of prior given parameter values indicating a communication problem influence the dialogue strategy (see DIALOGUE CONTROL).

TASK INTERPRETATION: This module has the control about the contents of the dialogue: it determines the sequence of the parameters needed for executing the caller's task, determines default values for parameters dependent on parameter values given so far by the caller, predicts parameters and parameter values realised possibly in the next caller utterance (this information can be used also for formulating alternative questions if only two or three values are possible, e.g. 'You work as an employee or free-lance?'), checks the consistency of the given (understood) information and informs the DIALOGUE CONTROL about inconsistencies.

All communication with the application system takes place here.

The module history contains all given data (only the final versions) during the whole dialogue, so that the dialogue can be finished at any time transferring all the data to either the application system (if this was not done during the dialogue) or to the terminal masks of the human operator.

GENERATION: The GENERATION module transforms an internal representation of the next system utterance into a sequence of words. Currently we use a direct transformation of dialogue goals and task parameters into a sequence of utterance templates, i.e. no real generation based on a semantic surface representation. This enables us to use pre-recorded and composed utterance parts (the templates) spoken by humans instead of synthesis. For more complex applications with a big or open vocabulary this will no longer be possible.

DIALOGUE CONTROL (cf. *Heisterkamp 1993*, *Heisterkamp & McGlashan 1996*): The DIALOGUE CONTROL is responsible for the overall control of the dialogue, i.e. it determines the next system initiative or reaction. This determination is based on the selection of one or more active goals to be achieved. Goals are added, their state is changed, or they succeed as a result of the interpretation of the messages sent by the different. So the CONTEXTUAL INTERPRETATION causes confirm goals (for confirming the understood information), whereas the TASK INTERPRETATION causes goals for requesting a parameter value, for informing about a found result or assumed defaults, and for explaining problems or inconsistencies.

Which goals are chosen for achievement is dependent on the actual dialogue strategy. The dialogue strategy allows a combined confirmation of several parameters with system initiative, a combined confirmation of several parameters without initiative, confirmation of only a single parameter in one system turn and yes/no question for confirmation or spelling mode for request. In all cases beside the first the system initiative (request for information) is realised alone in one system turn. The strategy is dynamic in that the system adjust the setting according to whether it encounters contradiction from the caller or not. If a contradiction occurs it is assumed that there were some problems with the understanding and the system tries to ask more and more specifically, finally degrading to yes/no questions and spelling mode. If the dialogue steps become again successful, the strategy is upgraded again.

4. Car Talk: Insurance dialogue in ACCeSS

In modelling dialogue structure, speech dialogue systems have so far largely relied on the assumption that they can map the task structure completely on the dialogue surface side. For the ACCeSS application this is not the case, due to non-disclosure of the complete task structure by the deployer of the system, as this concerns insurance tariff calculations and is thus classified information. In future applications, e.g. in the financial sector, this problem will become more common, as deployers of speech dialogue systems do not want company-confidential information be laid open to suppliers of speech technology, who may also serve their competitors.

Potential customers calling in to get a price offer have to provide those details of car type, insurance status etc. that are necessary to calculate a competitive premium rate to them. There are considerable logical dependencies between these different data. Some of them, like the congruency between car producer and type (e. g. Fiat Uno), the power of the engine etc. can be exploited in the dialogue to check for compatibilities and incompatibilities that indicate recognition or understanding errors. These dependencies are public knowledge. The inferences drawn from them are also used to request from the caller only the minimum information necessary, such as both to make the dialogue short (and thus to be able to handle more calls with a given number of lines) and to reduce the possibilities for misunderstanding and misrecognition.

Other dependencies, however, are not public knowledge. In fact, they concern the competitive position of that particular insurance company and thus cannot be disclosed to a technology provider who may also do business with other insurances. Some information received from a caller may influence the whole course of the dialogue in a manner that the technology provider may not be allowed to look into. The insurance may require that the dialogue be handed over to a human operator under certain conditions which are not and may not altogether be made explicit to or in the dialogue. Thus it is not possible to provide simple switches in the flow of the dialogue that would realise that behaviour.

These special non-disclosure requirements in the ACCeSS application do not allow to realise the task model as an integral part of the dialogue management component, as it is done for other applications. Here parts of this model are external and in the hands of the insurance company. We describe in the following how this mixed task structure is incorporated in the dialogue management to yield an overall flexible and coherent system behaviour.

The TASK INTERPRETATION module that manages the task model is dynamic in that it draws inferences from the state of information reached in the dialogue so far, and the overall model (see above). In the ACCeSS system, the task model and the inferences that go along with it have had to be divided: A part of the task model and of the inferences now are external, i.e. task knowledge like inferenced parameter values or the next required parameter are brought into the TASK INTERPRETATION module via the communication with the application system itself - that is, in out case, the insurance data base. So, task goals may be introduced form outside the TASK INTERPRETATION module as well.

As we are not allowed to give a real example we give a fictive one concerning only data that are publicly accessible. All cars are attached to one of up to 100 risk classes for liability, full body and partial body insurance (theft etc.). These risk classes are computed centrally from the damage and claim data of all German car insurers. So, the task of identifying the car in question is in fact a bi-layered mapping task. We have to get enough information to identify a unique set of classes. This may be easy, as e.g. all the Fiat Unos under 60 HP are in one class (comparable to the used car sale application shown in *Senneff, Polifroni & Zue 1996*). Suppose, however, the caller (and prospective customer) of the car insurance wants to insure a high risk car, e.g. a Ferrari Testarossa. Due to internal risk calculations, the insurance company may be willing to issue a full body insurance only if some requirements are met by the driver, like having reached a certain age, being married and thus driving (hopefully) more responsibly, etc. If the respective questions would be posed immediately after the car was identified as a Testarossa, this not only could scare off a potential customer for some other type of insurance, e. g. liability, but also might give hints as to tariff calculations to competitors, who also call in to check the premium rates.

We introduced the notion of „data groups“, such as car data, driver‘s personal data, insurance relevant data etc. (see fig. 1). These groups serve several purposes in structuring the dialogue both for the caller and the system. One purpose is that the part of the task model that is in the hands of the insurance company is allowed to introduce the request for the kind of information mentioned above, but tagged as belonging to one of these groups. For instance, if a full body insurance shall not be issued if the driver has not reached a certain age, even in this case the „normal“ logical and commonly expected dialogue is performed gathering first the car_data and the insurance data not laying open the internal preconditions of the insurance. Thus, the internal task model will generate the appropriate dialogue goal at the respective point in the dialogue. These points will neither be pre-determined nor laid open in the dialogue structure, but rather they will depend on some precedence factor attached to them by the insurance company.

In our system, the customer can determine to the desired extent what the structure of the dialogue will be, without having to either do the whole layout be himself or to disclose the complete rationale of his requirements to the dialogue system provider. In one respect, it is a step towards „self-organising dialogues“. In another, it is a step towards customer-oriented system design.

5. Results and Outlook

First evaluation of one subtask (determining the car type number out of a description of the car) shows for the speech recognition a word accuracy of about 80% using manageable graphs of word hypothesis (10 hypotheses per spoken word).

Currently we expand the application to its full functionality (see fig. 1). It is planned to install the first prototype at the end of the year at the insurance. This prototype will be known only to selected consumers. All dialogues will be registered and evaluated. The results of this evaluation (difficulties with the system behaviour or the system utterances, missing words and grammatical phrases, but also real failures) will be taken account for realising the final version of the system.

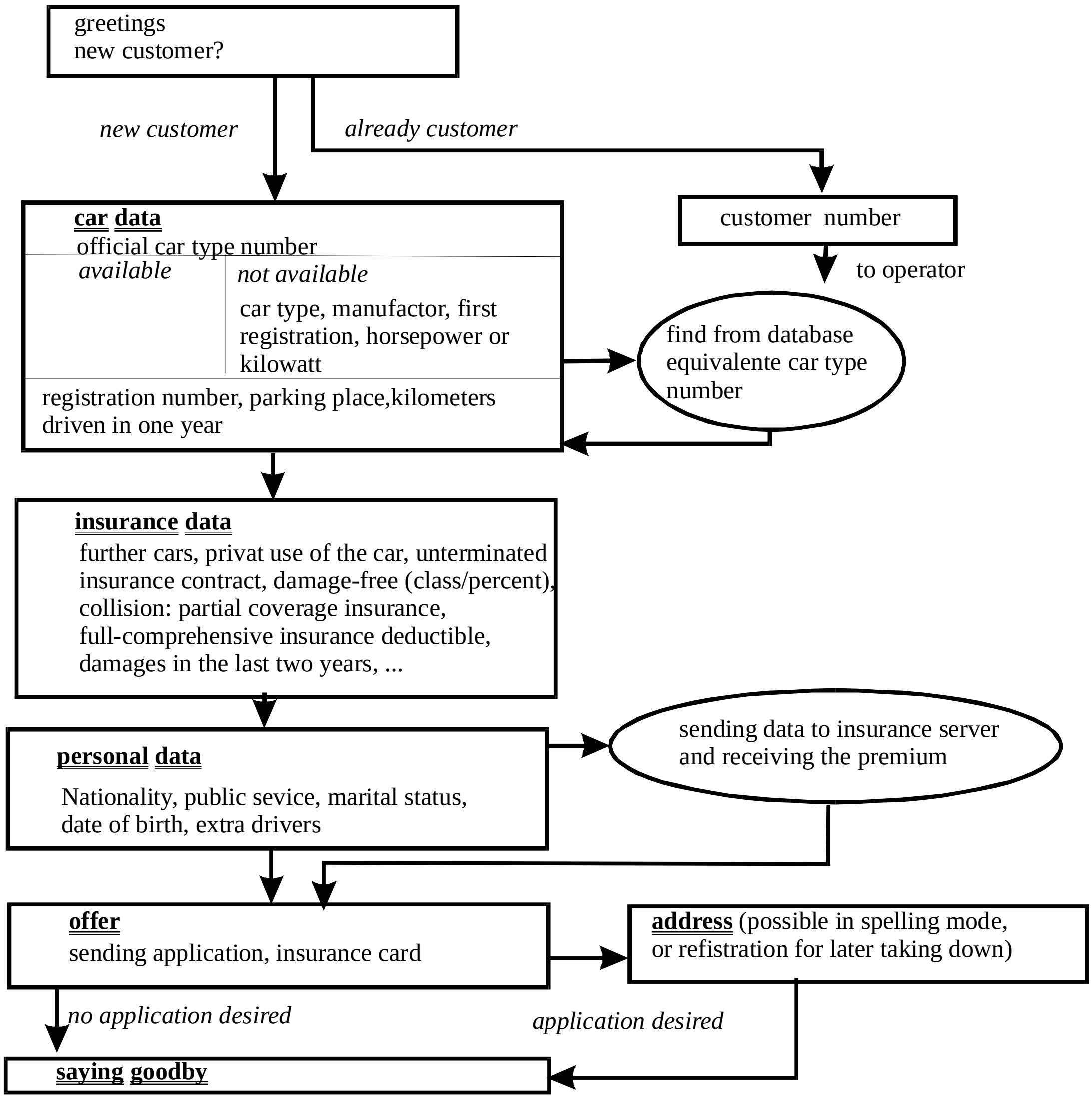

Fig. 1: Dialogue structure based on „data groups“

REFERENCES

A.Brietzmann, F.Class, U.Ehrlich, P.Heisterkamp, A.Kaltenmeier, K.Mecklenburg, P.Regel-Brietzmann, G.Hanrieder & W.Hiltl: (1994): Robust speech understanding. Proc. of ICSLP 94, Yokohama, 1994, p.967-970.

U.Ehrlich & G.Hanrieder (1996): Robust Speech Parsing. Workshop on Robust Parsing. ESSLLI '96. p.26-34.

P.Heisterkamp (1993): Ambiguity and uncertainty in spoken dialogue. Proc. of Eurospeech 93, Berlin, p.1657-1660.

P.Heisterkamp & S.McGlashan (1996): Units of dialogue management: An example. Proceedings of ICSLP 96, Philadelphia.

L.Hitzenberger & H.Kritzenberger (1989): Simulation Experiments and Prototyping of User Interfaces in a Multimodal Environment of an Information System. In: Tubach, J.P.; Mariani, J.J. (eds.) Eurospeech 89, European Conference on Speech Communication and Technology, Vol. 2, Paris. pp. 597-600

J.Krause & L.Hitzenberger (1992): Computer Talk. Hildesheim: Olms

S.Senneff, J.Polifroni & V.Zue (1996): Experiences with a field trial concerning used car sales by speech dialogue. Proceedings of ICSLP 96, Philadelphia.