An Integrated Dialogue System for the Automation of Call Centre Services

K. Georgila(1), A. Tsopanoglou(2), N. Fakotakis(1), G. Kokkinakis(1)

(1) Wire Communications Laboratory, Electrical and Computer Engineering Dept.,
University of Patras, 261 10 Rion, Patras, Greece
Tel:+30 61 991722, Fax:+30 61 991855, e-mail: {rgeorgil, fakotaki, gkokkin}@wcl.ee.upatras.gr

(2) KNOWLEDGE S. A., Human Machine Communication Dept.,
N. E. O. Patron-Athinon 37, 264 41 Patras, Greece
Tel:+30 61 452820, Fax:+30 61 453819, e-mail: atsopano@knowledge.gr

ABSTRACT

A human-machine dialogue system for the automation of call centre services is presented, which features two novel functions: a. It integrates spoken and written language processing through cooperation of a spoken dialogue component and an OCR which extracts handwritten information from application forms. b. It reads the forms and calls the applicants (instead of being called by them) in order to collect the missing information through a continuous speech, full sentence, machine-driven dialogue which allows the user to formulate his answers at will. The system has been developed for a Greek car insurance company in the framework of the EU-project: LE-1 1802, ACCeSS, but can be easily adapted to a different call centre application. The system is currently tested at the company's premises.

1. INTRODUCTION

In this paper we present a dialogue system developed in the framework of the EU-project "LE-1 1802, ACCeSS" [1] for a Greek car insurance company which participates as user in this project. The system integrates spoken and written language processing employing two components: An Optical Character Recognizer (OCR) and a Dialogue Component. Handwritten application forms filled out by prospective customers for the issuing of an insurance contract are processed by the OCR and the extracted information is stored on a provisional local database. The Dialogue Component, in turn, undertakes the collection of the missing information of the application forms by initiating calls to the customers. Considering that nearly 40% of the calls are answered by a third party, this function of the system highly complicates the dialogue.

Up to now the application forms are manually handled by typists, while the missing information is collected by human operators. Due to the high number of applications and the fact that almost 75% of them are incomplete, even a small automation of the whole procedure would yield considerable benefits for the company.

The steps performed by the system are as follows: Application forms with handwritten, isolated, upper case letters and numbers, entered either by fax or scanning are "read" by the OCR. The extracted optical information is stored on the local customers' database after confirmation/correction by a human operator. In case of missing data the Dialogue Component initiates a call to collect them provided it has been authorized by the operator. To this end, the system reads the stored information on the local database and prompts the customer to give the necessary answers to complete the blank fields. In case that the dialogue cannot be completed successfully, the call is transferred to a human operator along with the already gathered information. After completion of this procedure the blank fields of the customers' records are filled and a special module takes the responsibility to update the company's Central Customers' database with the collected information. In this way, it is ensured that the central database is protected from being contaminated by erroneous entries.

The paper is organized as follows: The system description is given in Section 2. Section 3 gives information on the hardware and software of the system. Results of evaluation experiments are given in section 4. Finally, a short summary and some conclusions are given in section 5.

2. SYSTEM DESCRIPTION

The system consists of two functional components: The OCR and the Dialogue. The Dialogue Component includes the following modules: The Telephone Line Interface, the Process Controller, the Dialogue Manager, the Speech Recognizer, the Semantic Interpreter and the Speech Generator. There are also a lexical database accessed by the Speech Recognizer, the Semantic Interpreter, the Speech Generator and the OCR and a customers' database, which connects the OCR with the Dialogue Component. The structure of the system is shown in Figure 1.

2.1. Telephone Line Interface

This module in cooperation with the Process Controller handles the PSTN calls. In particular it sets up the call to the customer who has to provide the missing information, it gives access of the incoming speech to the Speech Recognizer and directs the Speech Generator output to the telephone line.

2.2. Process Controller

The Process Controller is responsible for the initiation of different instances of the system's modules in cooperation with the dialogue manager in order to service multiple calls and handle several system resources. In cases where the input

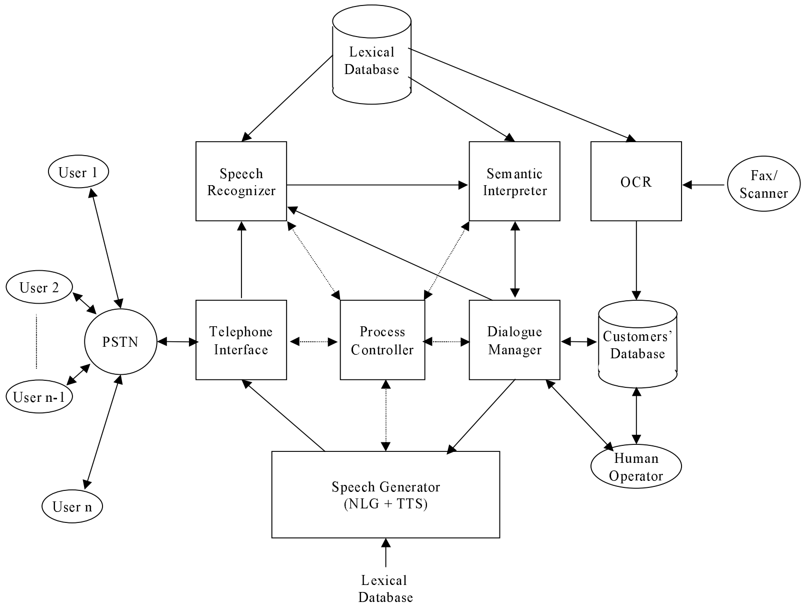

Figure 1: System's structure.

supplied by the user cannot be understood by the system even after clarification and/or confirmation questions, the Process Controller after notifying the Dialogue Manager directs the call to a Human Operator. The operator undertakes to resume the conversation with the user and recover possible errors, taking into consideration the history of the dialogue transaction and the already collected information.

2.3. Dialogue Manager

The Dialogue Manager controls the human machine interaction by following predefined dialogue scenarios. These scenarios have been formulated on the basis of a thorough analysis of both human-to-human dialogues and experimental results of WOZ dialogue simulation. In the first case 41 real dialogues recorded by the insurance company were transcribed. In addition to the sequence of words and turns in each identified dialogue theme, information about frequently occurring phenomena (pauses, pronounced hesitations, incomprehensible speech, sudden noise, etc.) was extracted as well as information about the call answering by the applicant or a third party. The information from the real dialogues was used to design the WOZ experiments which included 60 dialogues with males and females of different age, education and origin and contributed to the completion of the dialogue trees and the system prompts.

The Dialogue Manager on the basis of the output of the Semantic Interpreter at each dialogue node, forms a node-dependent structure which contains all the useful information supplied by the customer (i.e. the name of the function that must be executed at the current node, the variable names that keep the parameters given by the Semantic Interpreter etc). Then it queries/updates the local database, or prompts the user to give extra information or confirm the validity of the data provided at the previous node. It notifies the Speech Recognizer and the Semantic Interpreter on the dialogue expectations and the Speech Generator about the messages it has to provide.

2.4. Speech Recognizer

Two recognizers have been implemented and are currently tested. Initially a Speech Recognizer developed by WCL and Knowledge was used. This is based on a set of 808 basic units (phonemes and syllables i.e. two-phone, three-phone, four-phone and five-phone combinations of two or more consonants always ending in a vowel), described by a 5-state left to right continuous HMM. During the on-line recognition procedure the Frame Synchronous Viterbi Beam Search algorithm is used in order to produce the most probable sequences of basic recognition units taking into consideration the transition probabilities between them. Then the 10-best word sequences (hypotheses) are formed using the current lexicons and bigram probabilities between the words, and are passed to the Semantic Interpreter [2].

Recently a speech recognizer was built using Entropic's HTK toolkit. Each one of the 37 Greek monophones is described by a 5-state left to right continuous HMM, the parameters of which have been defined during the off-line embedded Baum-Welch re-estimation procedure. This set of monophone HMMs is used to create context-dependent triphone HMMs. This is done in two steps. Firstly, the monophone transcriptions are converted to triphone transcriptions and a set of triphone models are created by cloning all the monophones and re-estimating, which leads to a very large set of models, that is 6139, and relatively little training data for each model. Secondly, similar acoustic states of these triphones are tied to ensure that all state distributions can be robustly estimated [3].

The language modelling is based on word networks defined using HTK Standard Lattice Format (SLF). An SLF word network is a text file, which contains a list of nodes representing words and a list of arcs representing the transitions between words. A simplified example word network, which shows some possible answers to the question "What is the colour of the car?" is given in Figure 2. The node "COLOUR" points to a sub-lattice with the possible colours.

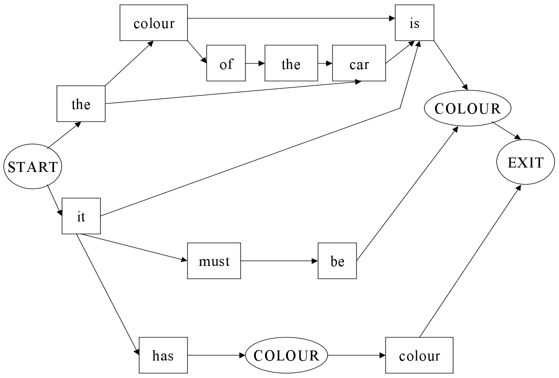

Figure 2: Word network for the colour of the car.

The HTK Application Programming Interface (HAPI) is used in the recognition phase. Given the set of the HMMs produced during the training stage, the SLF file for the particular dialogue stage and the corresponding dictionary, the HAPI decoder produces the 10-best paths of the word network and passes them to the Semantic Interpreter [4].

2.5. Semantic Interpreter

The linguistic analysis is based on Island Parsing, Pattern Matching and Frame-based representation techniques. The main knowledge sources are a Semantic Network and Frame-slot structures connected with each other. Simple bigram grammar rules are also used to assist the parsing process as well as to evaluate the recognition output [5].

Expectations that emerge from the current state of the dialogue model are used. These Dialogue Expectations are applied very early in the linguistic analysis so that pragmatic-irrelevant solutions are not considered in the first place. The results are structured appropriately and sent to the Dialogue Manager.

2.6. Speech Generator

The speech generator includes a Natural Language Generator (NLG) and a Text-To-Speech Synthesizer (TTS). The NLG constructs the system's response according to the dialogue design. It generates a natural language message, which is then converted into speech. The speech output of the system is a combination of prerecorded messages and synthesized speech. Prerecorded phrases are used for standard responses, speech synthesis for names, numbers etc., that cannot be prerecorded. A diphone, pitch synchronous TTS-synthesizer for Greek developed in our lab is employed. Alternatively, a previously developed formant TTS-synthesizer is tested.

2.7. Optical Character Recognizer (OCR)

The OCR converts an application form containing slots filled out with handwritten, separate, upper case letters and numbers, into a set of information, recognizable by the system. A preprocessor has the task to refit the form and extract the content of the slots. In order to do that, it compares the inserted document with the prototype form and rotates and shifts it to achieve the best match. Afterwards it undertakes the character extraction task for each slot of the refitted page. Each character is recognized by the OCR separately and the output is a sequence of characters. Then the information about the corresponding slot is used to activate the appropriate lexicon(s) (e.g. if the slot corresponds to the license number, it activates the number lexicon). By using this lexicon as input, and implementing the level building algorithm (dynamic programming), the most probable word(s) corresponding to the recognized character sequence is found [6].

2.8. Databases

The integrated system uses three databases: One containing the lexicons activated at each dialogue node by the system's modules, one local database containing personal information about the customers and the company's central database.

The lexical database consists of two correlated tables, one containing special fields necessary for the correct function of each module e.g., orthographic and phonetic transcriptions of words for the Speech Recognizer etc., and a second table filled with node specific information and constraints. The fields of the lexical database are independent of the application domain. By having one database for all the modules it is ensured that all the subparts of the system share the same lexicons. The customers' database contains application dependent fields and is updated by the Dialogue Manager according to the queries created by the information taken from the semantic interpreter. The Dialogue Manager provides a generic interface for any kind of SQL server commands embedded in C. The company's database includes the final and correct data of the clients. A special module has the responsibility to update this database with the collected information.

3. HARDWARE AND SOFTWARE

The Dialogue Component runs on a Pentium PC using an industry-standard telephone line interface card (Dialogic D41ESC). All the modules of the Dialogue Component are developed in C/C++ apart from the Semantic Interpreter, which is implemented in Prolog. The HAPI developed by Entropic in C/C++ is used for the speech recognition task. The OCR component is implemented in C/C++ as well and runs on a separate Pentium PC. The two PCs access the same local customers' database.

4. EVALUATION EXPERIMENTS

Evaluation of the performance of both the separate modules and the whole system has been continuously carried out during development and has guided the necessary modifications and improvements. In the following some results are presented from recent experiments by personnel of WCL.

Both speech recognizers have been trained with speech from the SPEECHDAT-II [7] telephone recordings, using 1000 male and female speakers, which cover a wide range of speech variation. Early tests carried out on our recognizer led to the decision to alternatively use the HTK recognizer. The tests performed up to now show that this approach produces better results. For the tests continuous speech by 100 speakers including 9752 words and 31628 triphones was used. The average percentages of accuracy, insertions, substitutions and deletions for context-dependent triphones and whole words, measured on this recognizer are shown in Table 1.

	A (%)	I (%)	S (%)	D (%)
Triphones	51.92	8.27	32.69	7.12
Words	72.84	2.94	20.63	3.59

Table 1: Average percentages of accuracy (A), insertions (I), substitutions (S) and deletions (D) for triphones and words (HTK recognizer).

To evaluate the Dialogue Component the dialogue nodes have been divided into five different categories, that is, nodes asking about information containing only numbers (postal code), combinations of letters and numbers (identity card's number), names of specific fields (city, car accessories), dates (birth date) and yes/no questions. The user is not limited as to how he will give the asked information. The system is able to handle full sentences of continuous speech.

	Once (%)	Twice (%)	3 times (%)	Failed (%)
Numbers	71.24/ 69.85	16.52/ 17.22	8.41/ 8.58	3.83/ 4.35
Letters/ Numbers	67.12/ 68.39	21.58 20.07	8.35/ 7.91	2.95/ 3.63
Names	78.04/ 75.94	14.81/ 18.11	5.39/ 4.26	1.76/ 1.69
Dates	63.71/ 65.04	21.15/ 20.32	10.75/ 11.23	4.39/ 3.41
Yes/No	93.58/ 91.82	5.51/ 7.08	0.48/ 0.71	0.43/ 0.39

Table 2: The results of the tests performed to measure the performance of the Dialogue Component (same dialogues / different dialogues on the same theme).

The performance of the Dialogue Component has been estimated counting the percentage of nodes that have correctly processed the customer's information immediately, they have been given the information twice, three times or they failed. After the information is processed three times unsuccessfully, the human operator undertakes the task to resume the conversation with the user. The tests were carried out by 28 speakers using the same dialogues and different dialogues on the same theme. The results of these tests are given in Table 2.

The OCR component has been initially trained and tested for English characters in order to assess the performance of the developed algorithms, due to the availability of large standardized databases for English [8]. The recognition accuracy for digits has been estimated to 90.21% using 50000 digits for training and 60000 for testing. For upper case characters 50000 samples (upper case characters) have been used for training and 11000 for testing. The recognition accuracy has been measured to 79.73%. For the training and testing of the OCR with Greek characters completed handwritten forms supplied by the insurance company were used. Due to the smaller number of training data the measured recognition rate is lower than in English (87.39% for digits, 72.94% for upper case characters).

Recently the integrated system has been installed at the premises of the car insurance company and is tested by personnel of this company. The results are regularly communicated to the developers and serve as basis for improvements. After this stage which will last for six months, the system will be given to customers' service. A formal evaluation lasting for three months will then be carried out.

5. SUMMARY AND CONCLUSIONS

An Integrated Dialogue System combining spoken and written language has been presented. It consists of an OCR and a Dialogue Component. Information extracted from application forms by the OCR component updates a local database. The Dialogue Component takes the responsibility to call the customer in order to gather data for the blank or ambiguous fields of the local customers' database. When all the fields have been completed and checked by a human operator, the company's database is updated. The system is currently running at a car insurance company and tested by its personnel. The results are used for further improvements. A formal evaluation of the system will be carried out using real customer calls.

6. REFERENCES

1. ACCeSS, "*Automated Call Centre Through Speech Understanding System*", LE-1 1802.

2. K. Georgila, A. Tsopanoglou, N. Fakotakis and G. Kokkinakis, "*A Dialogue System for Telephone-based Services Integrating Spoken and Written Language*", IVTTA '98, 4th IEEE workshop on Interactive Voice Technology for Telecommunications Applications, Torino, Italy, Sept. 29-30, 1998 (to be presented).

3. S. Young, J. Odell, D. Ollason, V. Valtchev and P. Woodland, "*The HTK Book*", Entropic Cambridge Research Laboratory, March 1997.

4. J. Odell, D. Ollason, V. Valtchev and D. Whitehouse, "*The HAPI Book, A description of the HTK Application Programming Interface*", Entropic Cambridge Research Laboratory, December 1997.

5. A. Thanopoulos, N. Fakotakis and G. Kokkinakis, "*Linguistic Processor for a Spoken Dialogue System based on Island Parsing Techniques*", Eurospeech '97, 5th European Conference on Speech Communication and Technology, Rodos, Greece, Sept. 22-25, 1997, Vol. 4, pp. 2259-2262.

6. N. Liolios, N. Fakotakis and G. Kokkinakis, "*Form Identification and Skew Detection from Projections*", EUSIPCO '98, 9th European Signal Processing Conference, Rodos, Greece, Sept. 8-11, 1998 (to be presented).

7. I. Chatzi, N. Fakotakis and G. Kokkinakis, "*Greek speech database for creation of voice driven teleservices*", Eurospeech '97, 5th European Conference on Speech Communication and Technology, Rodos, Greece, Sept. 22-25, 1997, Vol. 4, pp. 1755-1758.

8. USPS Office of Advanced Technology, "*Database of Handwritten Cities, States, ZIP codes, Digits, and Alphabetic Characters*".