STACC: AN AUTOMATIC SERVICE FOR INFORMATION ACCESS USING CONTINUOUS SPEECH RECOGNITION THROUGH TELEPHONE LINE

Antonio J. Rubio, Pedro García, Ángel de la Torre, José C. Segura, Jesús Díaz-Verdejo,
María C. Benítez, Victoria Sánchez, Antonio M. Peinado, Juan M. López-Soler, José L. Pérez-Córdoba
Dpto. de Electrónica y Tecnología de Computadores
Universidad de Granada, 18071 GRANADA (Spain)
e-mail rubio@hal.ugr.es telf: 34-58-243193 fax 34-58-243230

ABSTRACT

This work presents the STACC, *Sistema Telefónico Automático de Consulta de Calificaciones* (Automatic Telephone System for Consulting Marks). This system has been developed at our laboratory during 1996 and implements a service through telephone line that allows the students to consult by speech their marks after the exams by means of a simple phone call. This experience provided us an interesting point of view about the problems of real applications of speech technology. In this work we describe the system and some statistics about the use of STACC by the students are presented.

1. INTRODUCTION

During the recent years, an important progress in the spoken language technologies has been carried out. In the field of Automatic Speech Recognition (ASR), system accuracy is being improved for tasks that used to be unapproachable and under reasonable restrictions the performance is satisfactory [1].

However, when speech technology is applied out of the laboratory conditions, some new difficulties appear. In order to develop a real service that makes use of the ASR technologies it is necessary to focus the problem from a new point of view:

- The acoustic conditions are very different to the ones of the laboratory. The system must be robust against noise or the incorrect use on the part of the customer.
- The system has to work in real time.
- The main objective is to obtain a system acceptable by the customer. The system must be easy to use and the information interchange between the customer and the system must be fluid. If the system is not user-friendly, after a "fever of the new automatic system" it is left and forgotten by the customer. Systems that accept continuous speech allow a more fluid dialog and are more easily accepted.
- Of course, the system must be reasonablely efficient to be accepted by the users.
- Unexpected problems are always found.

Recently, an important effort has been done to develop ASR systems to deal with real tasks, like *dictation systems* [2], or *telephone applications for switchboard* [3] [4]. Thanks to the progress in the speech technologies, and due to the wide range of possible applications, the effort in the design and implementation of dialog systems through telephone line is being increased [5] [6].

In this work, we present a telephone application of speaker-independent continuous speech recognition technology. This application offers an information service in which the customer accesses to the information through telephone line by speech. By means of a dialog driven by the system, the customer can access to the information available in a database. STACC allows the students to consult their marks by mean of a simple telephone call. The main requests of the service are:

1. *confidentiality*: the information is assumed to be private,
2. *security*: it is not acceptable that the system provides a wrong mark, and
3. *effectivity*: the success of the implementation is to make easier the mark consulting process. Thus, an important criterion for the evaluation of the system is the acceptation on the part of the users.

The development of STACC aims three main objectives: (1) to analyze the problems we found when the speech and language technologies are applied to a real situation, (2) to create an environment where we can perform real tests of the ASR methods and (3) to provide a real service that make use of the ASR technology. In this work, the STACC is presented and described, and some statistics about the usage of the system by the student are discussed.

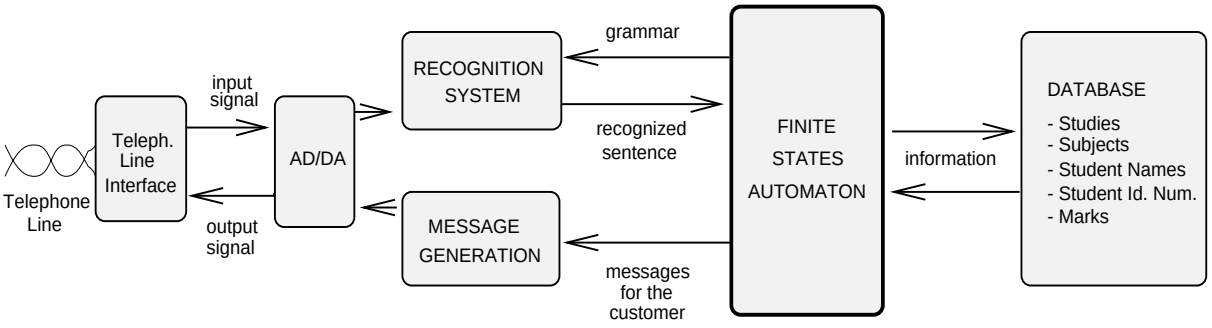

Figure 1: *Block diagram of STACC*

2. DESCRIPTION OF STACC

2.1. Elements of STACC

Figure 1 shows the block diagram of the STACC system. STACC is composed of the following elements: A *finite-states automaton* that controls the task flow; the *telephone line interface* and the *AD/DA converter* to handle the input/output speech signal of the system and a *database* containing the information that the users can consult: about the university courses, the subjects for which the marks are available, and, for each subject, a list of students, their identification numbers [1] and the marks.

Depending on the state, the flow control automaton obtains the particular grammar (lexicon and rules) to be used by the recognition system, sends to the message generation block the output information and reads from the recognition system or the database.

STACC is working on a *SUN-ELC* workstation. The AD/DA converter is integrated in this computer. The only specific hardware not integrated in the workstation is the telephone line interface. This interface interchanges control information by the serial port RS-232, and amplifies and adapts the input and output levels. The input signal is filtered, sampled (f_s=8KHz) and quantized (μ-law, 8 bits) by the AD/DA converter.

The recognition system includes a voice activity detector block. The signal is pre-emphasized and segmented into frames of 30ms; the frame period is 10ms. Every frame is represented by a feature vector containing 14 Mel Frequency Cepstral Coefficients, the energy, and the first and second derivatives. The acoustic units are *context independent Phone Like Units* (PLUs) which are modelled by *Semi Continuous Hidden Markov Models*. The language model is a *regular deterministic grammar*. The grammar (vocabulary and rules) is different depending on the state of the automaton and is automatically loaded in a dynamic way. The typical size of the vocabulary is around 300 words. The recognition system provides the best sentence allowed by the current grammar by using the Viterbi Beam Search algorithm. In order to adapt the system to noise conditions and to speaker-independent recognition, the acoustic models have been trained with sentences uttered by 200 speakers through telephone line.

Since the set of messages to be send to the customer is small, the message generation block has not been implemented by a *text-to-speech* converter. The set of messages were previously recorded, sampled and stored, and are played according to the automaton state.

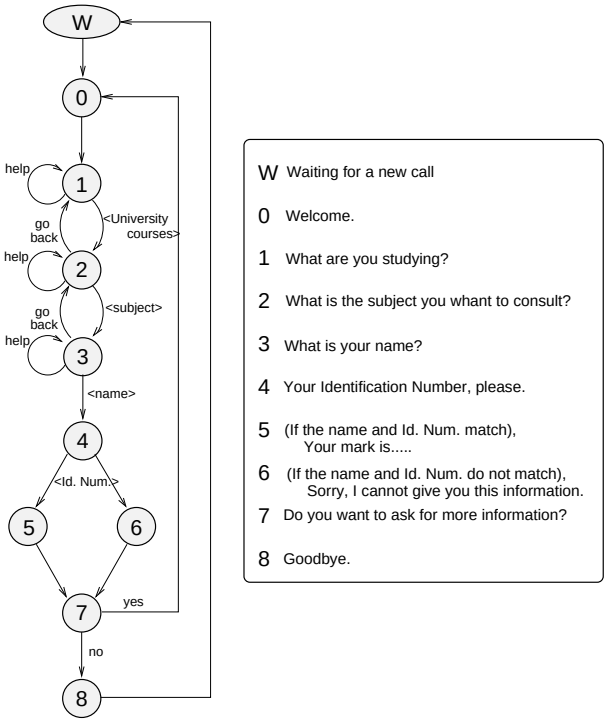

Figure 2: *Task flow*

2.2. Task flow

Figure 2 shows the task flow of STACC. Initially, the system is waiting for a telephone call (W). When the telephone line interface detects a call, the line is hanged up and the system sends a welcome message and provides very brief instructions (0). Then, the system asks for the university courses to be consulted, and waits for the student answer (1). The student informs about the studies he is coursing. Then, the system asks for the subject to be consulted

[1] In Spain, there is an official Id. document with an 8 digits identification number.

Table 1: *Some statistics about STACC*

	Feb 96	Jun 96	global
Number of University courses	2	2	-
Number of available subjects	4	10	-
Average number of students per subject	65.3	72.1	-
Total number of possible requests	261	721	-
Considered calls	215	144	359
Mark requests	248	156	404
Consultations successfully attempted	209	146	355
Efficiency of the service	84.3%	93.6%	87.9%
Number of wrong marks	0	0	0
Rate of wrong marks	0%	0%	0%
Number of sentences processed	1191	785	1976
Sentences correctly recognized	1021	703	1724
Sentence-recognition rate	85.7%	89.6%	87.5%
Female customers	14.8%	23.6%	18.3%
Male customers	89.2%	76.4%	81.7%

(2), the name (3) and the identification number (4). The system compares the recognized name and the recognized identification number. If they match, the system informs about the mark (5). Otherwise, it is assumed that there is a misrecognition or the customer is not authorized to know the mark. Then the system does not inform. The system asks if the student wants to consult more marks (7). If he does, the system starts from state (0). Otherwise, the system sends a goodbye message (8) and goes to the waiting state (W) after hanging down the line.

An example of a conversation between the system and the customer could be the following:

-system- *Welcome to the automatic telephone system for consulting marks. If you need to know the instructions of use, say "help". If not, say the name of the studies you want to consult.*

-caller- *Electronics Engineering.*

-system- *Say a subject of Electronics Engineering.*

-caller- *Electronics for Communications.*

-system- *You have selected Electronics for Communications. What is your name and family name?*

-caller- *Jose Campillo Baca.*

-system- *Say the code of your identity card, number by number. For example, "two four one five seven two nine two".*

-caller- *Two five one seven five four eight nine.*

-system- *According to the available information, your grade is "notable"* [2]*. Do you want to consult more marks?*

-caller- *No.*

-system- *Good-bye and thanks for using our system.*

In the STACC system, the extraction of information from the recognized strings has been simplified (in contrast to other dialogue systems). Therefore, for cooperative customers, the success of the STACC only depends on the efficiency of the recognition block.

[2] In Spain, the grade "notable" corresponds to 7-8 in a scale of 10

The possibilities offered to the customer at each state are relatively limited and a spontaneous dialogue is not allowed by the system: the dialogue is driven by the *finite states automaton* and not by the customer. This allows the use of very restrictive production rules in the recognition system and to use a reduced vocabulary at each state, which makes easier the recognition task, and improves the performance of the recognition block, and therefore, the efficiency of the STACC.

3. SOME STATISTICS ABOUT USAGE OF STACC

STACC was developed at the beginning of 1996. At the end of February 1996 it was working in a preliminary stage [3]. At this stage, only a few number of subjects were available, and students were warned against possible errors (like wrong marks or abnormal dialog flow....) and the behavior of the system was supervised. After a month (when the peak of consultations was passed), we analyzed the results and corrected some problems of STACC. The next peak of consultations was in June 1996. In this stage, the system unsupervisely provided the marks to the students of a greater number of subjects.

Table 1 shows some data obtained from the analysis of the calls. In order to obtain these statistics, the audio signal was recorded and the speech was listened and contrasted to the information provided by the automaton. In this analysis, only the calls performed by students are considered (calls for checking STACC are discarded), and only in the case of being the requested subject available (between the time when the exams are written and the marks are available, there are a lot of calls which could not reach their objective).

A wrong mark has never been given to any student. The rate of successful calls is greater than 85%. The effi-

[3] At Spanish University, most of the exams are made in February, June and September.

ciency of the service and the recognition rate was increased in June. If the call is not successfully ended, we have observed that the student usually obtained the information in a second call, and no student required more than three calls.

One important difference between this two periods is that the dialog was more fluid in June, because of the improvement of the voice activity detector. In the preliminary stage, the problems with activity detection made the students repeat a lot of times some data, specially if they spoke from a very noisy environment. Even though the efficiency of the recognizer was high in February, the dialog was not fluid, and the consultation of marks took a lot of time, because a high number of sentences was not processed by the speech recognizer and the students had to repeat the sentences. This fact reduced the confidence of the students on the information provided by STACC, and they usually confirmed the marks with several calls. After the improvement of the voice activity detector and the acoustic models, the dialog was significantly more fluid, the consultations were faster and the information provided by the system was easily accepted. This explains the reduction of the number of calls observed from February to June.

Two factors are involved in the improvement of the service observed between February and June. Firstly, the STACC was adapted to the work conditions, and the main problems observed were solved. On the other hand, the service was improved because of the adaptation of the students to the STACC: the students learned how to obtain the information.

In September 1996, STACC included 11 subjects (from 2 university courses), containing a total of 695 available marks, and 272 calls were processed. In this period, the efficiency of the consultation service was 94.5%. There are no statistics about this period because the speech was not recorded at this stage. Currently, STACC provides the marks of the February exams. In this period, the number of marks in the database is 227. The system processed 170 calls and 89.4% of them were successfully processed. The observed reduction in the rate of successful calls is due to the fact that these students were no experienced in the use of the STACC (while most of the student who used it in June and September were experienced). In the next months, we hope to extend the service to a greater number of subjects.

4. CONCLUSIONS

STACC is an experimental system and some additions are under way (like reducing the processing time, robustness against noise, extension of the offered services, etc.). However, the main requests of STACC (confidentiality, security and effectivity) were satisfied and the system provides a real service that makes easier the operation of consulting marks (especially for the students who live far away from the University center). Our experience constitutes an example about how the application of spoken language technologies to real services will be a reality in a short time. The development of the computer technologies, the improvement of the speech algorithms and the effort focused onto the spoken language dialogue systems, make reasonable to expect, for a not very remote future, the generalized use of voice controlled systems.

5. ACKNOWLEDGEMENTS

The authors would like to thank Mr. A. Balogh from the Technical University of Cluj-Napoca, Romania, for his help in the design and implementation of the telephone line interface for the STACC.

This work is funded by CICyT (Spanish Governmental Research Agency) under project TIC96-0956-C04-04.

6. REFERENCES

[1] L. R. Rabiner, "The impact of voice processing on modern telecommunications," *Speech Communication*, vol. 17, pp. 217–234, Nov. 1995.

[2] V. Steinbiss, H. Ney, U. Essen, B. H. Tran, X. Aubert, C. Dugast, R. Kneser, H. G. Meier, M. Oerder, R. Haeb-Umbach, D. Geller, W. Hollerbauer, and H. Bartoisk, "Continuous speech dictation - From theory to practice," *Speech Communication*, vol. 17, pp. 19–38, Aug. 1995.

[3] G. J. Vysotsky, "Voice Dialing - The first speech recognition basded service delivered to customer's home from the telephone network," *Speech Communication*, vol. 17, pp. 235–247, Nov. 1995.

[4] G. Riccardi, A. L. Gorin, A. Ljolje, and M. Riley, "A spoken language system for automated call routing," in *Proc. of ICASSP-97*, pp. 1143–1146, 1997.

[5] N. M. Fraser and J. H. S. Thornton, "VOCALIST: A robust , portable spoken language dialogue system for telephone applications," in *Proc. of Eurospeech-95*, pp. 1947–1950, 1995.

[6] D. Albesano, P. Baggia, M. Danieli, R. Gemello, E. Gerbino, and C. Rullent, "DIALOGOS: A robust system for human-machine spoken dialogue on the telephone," in *Proc. of ICASSP-97*, pp. 1147–1150, 1997.