IDAS : INTERACTIVE DIRECTORY ASSISTANCE SERVICE

G. Lehtinen[1], S. Safra[1], M. Gauger[1], J.-L. Cochard[2], B. Kaspar[3], M. Hennecke[4],
J.M. Pardo[5], R. Córdoba[5], R. San-Segundo[5], A. Tsopanoglou[6], D. Louloudis[7], M. Mantakas[8]

[1]Swiss Federal Institute of Technology Zurich, Switzerland
[2]Swisscom AG, Switzerland
[3]T-Nova Berkom, Germany
[4]TEMIC Telefunken microelectronic GmbH, Ulm, Germany
[5]Grupo de Tecnología del Habla. Universidad Politécnica de Madrid, Spain
[6]KNOWLEDGE S.A. (DIS Group), Language Engineering Dept., Patras, Greece
[7]Wire Communications Lab., University of Patras, Greece
[8]Hellenic Telecommunications Organization (OTE), Athens, Greece

ABSTRACT

In the EU funded IDAS project demonstrators providing a partially automated interactive telephone-based directory assistance service are being developed by ten partners from Germany, Greece, Spain and Switzerland. A first phase was completed with limited prototypes built, tested and demonstrated, showing the feasibility of such an application. In a second phase improved demonstrators with higher directory coverage and more human-like dialogues resulting in a higher automation rate are being built which will be integrated into small-sized systems operating in real life.

1. INTRODUCTION

The IDAS (Interactive telephone-based Directory Assistance Services) project addresses the challenging problem of automating the provision of directory assistance services to the public over the telephone network. While complete automation of the directory assistance service is not within reach during this project, the project will realize demonstration systems for German, Greek, Spanish and Swiss telecommunications operators which will provide a partially automated service, and move from experimental solutions up to small-sized systems operating in real life.

The technical challenge that has to be tackled makes high demands on each of the speech processing components: a speech recognizer that can distinguish the uttered word out of a large vocabulary, independently of the speaker's voice and the (mostly poor) signal quality, a speech production system able to speak out any imaginable phone directory entry (containing names and words from different languages), a dialogue component that can interpret user inputs and to ask the right questions in order to guide the users quickly to their desired information and out of misunderstandings.

However, the challenges in the IDAS project go beyond this technical level. The success of the project depends also on whether the demonstrators are able to meet the telecommunications operators' demands — that the costs of directory assistance services can be significantly reduced while still ensuring customer satisfaction and upholding the company's reputation. This, in turn, depends mainly on whether the end user's expectations can be met in terms of user-friendliness, speed, and reliability.

It is important to achieve user satisfaction from the beginning, once a system is operating in real life. However, speech technology is still far from being perfect. One of the project's main focuses is therefore to provide the user with a high success rate, independently of the current state of technology. To this end, the system design incorporates a so-called operator fallback component as will be explained in the next section.

A partially automated directory assistance system evolving from IDAS that is able to process at least the "routine" inquiries could offer a quick, cheap and highly available service to the public, thus reducing service providers' costs and increasing their competitiveness at the same time. But once it proves itself, the same technology can also be used to build many new, flexible services offering per-customer and up-to-the-minute information.

2. SYSTEM OVERVIEW

Special emphasis was placed on fostering synergy between the partners. To that end, a common system architecture with common module APIs was agreed upon, allowing the exchange of modules between the systems and thus making the IDAS system easily adaptable to new environments and even to new applications.
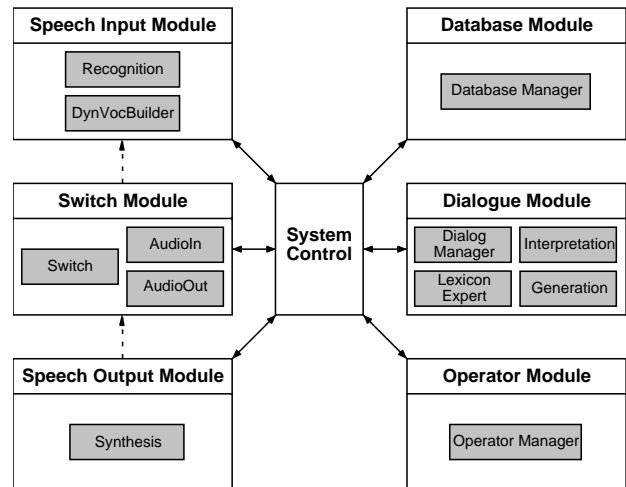

Figure 1: Common IDAS system architecture

Figure 1 shows the common system architecture containing

the following modules:

- The **Switch Module** offers an interface to the line and provides switching functionalities. Input and output audio streams flow directly between the Switch Module's AudioIn and AudioOut components and the Speech Input Module or the Speech Output Module, respectively.

- The **Speech Output Module** is responsible for the generation of speech signals either by synthesizing speech from text or by playing prerecorded speech files.

- The **Speech Input Module** handles speech recognition. An important requirement for user friendliness is to allow mixed initiative dialogues in contrast to pure system guided dialogues. Therefore the Speech Input Module must be able to process not only single word utterances but also whole phrases. All phrase structures expected at each dialogue step (dialogue context) have to be described in a grammar or language model. For each single recognition the appropriate context can be set, i.e. the active grammar can be chosen and the vocabulary can be dynamically generated depending on the dialogue history and/or database results.

- The **Database Module** encapsulates the given electronic telephone directory of each country and other locally stored or online information sources and offers an abstract interface to them.

- The **Operator Module**'s responsibility is to transfer all information gathered so far to a human operator in case that a problem occurred in the dialogue which the system recognizes as going beyond its capability to recover from. The operator has the option to listen to the customer utterances recorded by the system and – if possible – to complete the task without the customer being aware of any human interaction. Or the operator can connect to the customer and complete the task in direct customer interaction.

- The **System Control Module** integrates the whole system by drawing from the functionalities of the other modules and managing the data transfer between them.

- The **Dialogue Module** is consulted by the System Control Module for every dialogue step and is thus the "brain" of the whole system. A dialogue step is a conversational unit containing a prompt and a recognition phase or a database request. The Dialogue Module is responsible for composing these dialogue steps, i.e. generating user prompts and database requests, for interpreting recognition and database results and for setting dialogue-step dependent parameters for all modules. It also triggers a transfer to the operator in case of troubles. The lexicon expert expands the recognition hypotheses with respect to homophones and expands abbreviated words to the full form words. The semantic interpretation component carries out a linguistic analysis and assigns meaning to the word or phrase according to the semantic context.

All modules except the Dialogue Module are event driven, i.e. work asynchronously allowing a sophisticated control flow including time-constrained interaction and a barge-in mechanism. Of course, a barge-in mechanism requires echo-cancellation capabilities included in the Switch module.

3. THE IDAS DEMONSTRATORS

For the IDAS project, two demonstrators were planned, the first (D1.1) being more system-driven, showing the general feasibility of the project, the second (D1.2) allowing more natural user input and being more oriented towards an actual product.

To accommodate the differing needs of the telecommunication companies, such as the country's language or the environment into which the system has to be integrated, each demonstrator was built in four different versions, one for each country.

The partners of IDAS implemented and evaluated the first prototypes in 1999 which were demonstrated at the European Commissions project review in Luxembourg. Currently the more sophisticated and powerful demonstrators are worked on which will be able to cover a large vocabulary and conduct more human-like dialogues.

3.1. The Swiss Demonstrators

Both Swiss demonstrators are being developed in collaboration with TIK of ETH and CT-SPI of Swisscom AG.

Compared to the demonstrator D1.2 the D1.1 was restricted with respect to the following points:

- Restricted data base with some 117000 entries
- Vocabulary restricted to 100 place names, 100 last names and 100 first names.
- Only private entries
- Simple system driven dialogue
- No operator fallback used during the field tests
- No barge-in due to the lack of an appropriate echo-cancellation

After completing the system a preliminary test was run with IDAS co-workers to eliminate obvious system faults and fine-tune the dialogue as well as to verify the the experimental setup itself and the proper functioning of the assessment tools.

The actual field test was performed within 3 weeks with some thirty participants, resulting in an overall of about 500 test sessions. In each session a test person had to solve a randomly selected telephone directory inquiry task.

Apart from the quantitative measurements the test persons were also interviewed at the end of field tests to assess some important subjective criteria.

The following table shows the WER values for each of the main utterance types "Place", "Name", "First Name", and answers to "Yes/No" questions.

	WER_{in}	OOV	WER_{total}
Place	9.8%	0%	20.4%
Last Name	19.5%	0%	31.9%
First Name	9.6%	0%	16.6%
Yes/No	3.8%	60%	6.4%

The count is based on manually annotated data collected within this field-test. Two WER are given, a first one considering only utterance segments containing words of the active vocabulary (WER_{in}), and a second one, taking into account all utterances (WER_{total}). The high discrepancy is due to pre-recognition problems (bad start/endpoint detection), a poor Out-Of-Vocabulary model and poor phonetic transcriptions in the lexicon (automatically generated with a speech synthesis system).

The following overall task completion rates were achieved:

Successful	77.54%
Unsuccessful	20.32%
System error	2.14%

Despite the rather low recognition results even the simple dialogue used in demonstrator D1.1 could recover many of the errors. In the subjective evaluation the system achieved a satisfactory overall acceptance.

The second demonstrator currently being developed will improve several of its predecessors disappointing flaws:

- Better OOV-modeling
- Better Start/Endpoint detection
- The simple state-machine based dialogue will be replaced by the more sophisticated dialogue engine from our German partners allowing the development of more complex and subtle dialogue scripts
- Handcrafted phonetic transcriptions of the vocabulary

Moreover the vocabulary size will be extended from 100 to about 5000 words of each category, resulting in considerably higher coverage of the Swiss telephone directory.

Also the new demonstrator will incorporate TIK's new Polyglot Speech Synthesis and thus be able to articulate place, street and person names more clearly. This is of special importance in Switzerland having four official languages.

3.2. The German Demonstrators

Both German demonstrators are being developed in a collaboration of TEMIC Telefunken microelectronic GmbH, DaimlerChrysler AG and T-Nova Berkom and are installed at a call center of the Deutsche Telekom AG.

The goal of the German system was to provide directory assistance for all of Germany from the start. Thus, the database is not restricted but contains the full German directory. The recognizer vocabulary however is limited to only the most frequently requested names. In Demonstrator D1.1 there are 5000 city names, 10000 surnames and 5000 first names. Only a few company names were included to test that part of the dialogue. Nevertheless, the operator fallback still allows the system to find any entry in the telephone book.

Unfortunately, installation of the German system at the call center was delayed due to reasons outside of our control. For this reason, testing was limited to in-house tests at first. The actual field tests and evaluation are still on-going, but preliminary results look very promising. Final results will be available well before the conference.

In the following, we list some problems encountered which affect the success of the system.

- As it was to be suspected, some entries in the directory assistance database are not consistent. Among other effects, some information appears in the wrong place (e. g. title within the slot provided for first names).
- It appears that our name lexica need improvement. Some names are transcribed poorly, yielding in both poor recognition and poor speech synthesis. Furthermore, we encountered word pairs with seemingly irrelevant differences in orthography, but different transcription. This may lead to peculiar recognition results, as the lexicon does not offer a concept for phonetic similarity.
- Finally, the dialogue is often affected by poor perception of names as synthesized in confirmation questions. This is not only due to poor transcription, but rather to weaknesses of the synthesizer employed.

The second demonstrator will benefit from the experiences gained with the first demonstrator in several respects:

- vocabulary sizes: The number of names is increased to 50000 city, 100000 last and 10000 first names.
- In addition, 10000 company and 10000 street names will be included.
- The system correctly handles entries with more than one number.
- A number of improvements to the dialogue.

With these changes, the coverage of the German directory is greatly increased. It is expected that at least 75% of all calls can be handled automatically, whereas the operator time on the remaining 25% is reduced by about half.

3.3. The Greek Demonstrators

The Greek demonstrators were developed by KNOWLEDGE S.A. and WCL with the contribution of the Hellenic Telecommunications Organization (OTE). The first Demonstrator constitutes the baseline Directory Assistant and the second Demonstrator is an improved version with a larger vocabulary and much new functionality.

The D1.1 Demonstrator has the following characteristics:

- Very large recognition vocabulary (10000 surnames, 350 first names and 350 city/town names)
- The database consists of more than 5.000.000 entries and it is similar to the functional Greek directory database of OTE.
- It accepts only private telephone queries
- Implicit confirmation for the place name and the first name is supported.
- Fully synthesized output (TTS is provided by WCL)
- Multi-channel application
- Operator fallback has been included (the recorded data and the recognizer's results are passed to the operator)

The system was evaluated during a three-step evaluation procedure: in-lab glass-box, controlled black box and field black-box field. During the first phase, the system was evaluated by 100 employees of KNOWLEDGE and WCL.

After the in-lab tests, the D1.1 was deployed at OTE premises in Patras where the controlled and field black-box evaluation were carried out. The 200 callers followed a specific procedure that was described in questionnaires that were circulated by WCL.

The results of the three phases are summarized on the following table showing the recognition accuracy per task.

	1st phase	2nd phase	3rd phase
Call Type	100%	100%	100%
Place	95.5%	93.5%	89.6%
First name	97.6%	94.2%	91.1%
Surname	73.1%	68.6%	63.0%

Taking into consideration that the effectiveness of a spoken dialogue system can not be measured just by the recognition accuracy we also estimated the transaction completion rate and the mean task duration. We found that more than 53% of the calls were completed correctly by the system. The operator fallback option was enabled for about 42% of the calls. About 5% of the incoming calls were either cancelled by the users or the caller got mistaken results. The mean task duration was measured to be almost 46 seconds (multiple trials of the callers to correct the query are included).

The in-lab tests of the second demonstrator have been completed and the system will be installed at OTE premises by

February, 2000. The D1.2 has been improved in many points compared to the D1.1. The most significant differences between the two systems are:

- Vocabulary contents (88.000 last and 2500 place names)
- Barge-in
- Implicit confirmation for every task
- Overanswering
- Handling of different names with equal pronunciation (same phonetic but different orthographic transcription)
- A semi-spelling mode has been added where the caller is prompted to utter the first 3 letters of the surname of the required person
- The dialogue has been redesigned so that it produces a more user-friendly and flexible system.

3.4. The Spanish Demonstrators

The Spanish Demonstrator is being developed completely in the Grupo de Tecnología del Habla at the Universidad Politécnica de Madrid (UPM). The role of UPM in the project is to expand IDAS services to a new language, Spanish, solving initial problems for this language and service. Due to budgetary restrictions, it was not possible to include two phases in the demonstrator development (demonstrator 1 and 2) nor a telecommunication operator for integration.

By the time of the project review in February 99 we had a first version of the demonstrator ready, using context-independent models. The demonstrator had the following characteristics:

- Representative database with about 1 million entries.
- 4 different vocabularies: cities, first-names, surnames and company names, all of them using 1000 words.
- Provides phone numbers for private users and companies.
- A simple system-driven dialogue.
- No operator fallback used during the field tests.
- All the system messages were generated using our Spanish text-to-speech system [6].

The field tests were performed within 4 weeks. 33 participants (21 male and 12 female) were asked each to obtain 20 telephone numbers (10 for private entries and 10 for companies), giving a total of 660 sessions. The following table shows the error rates (WER) for each vocabulary (using either the best hypothesis alone in the dialogue or the next best as well):

Candidates	Cities	Companies	Last names	First names	Average
1	24.1%	14.4%	33.3%	32.1%	26.0%
2	16.8%	9.8%	25.5%	22.9%	18.7%

The error rate for the companies is lower. The reason is the longer average duration for company names (14.3 phonemes per word) which allows a better discrimination. The worst error rate corresponds to first names and surnames, as they have the shorter average duration (7.2 and 6.2 phonemes, respectively). Another important factor is the confusability between words belonging to the same vocabulary. With a dynamic programming algorithm we computed the phonetic distance between every word and the closest one in the vocabulary. We obtained the following average distances: surnames 2.0, first names 2.3, cities 3.3 and companies 6.8. As we expected, the worst error rates correspond to the vocabularies with smaller average phonetic distance.

We are right now finishing the development of the second version of our demonstrator. We will expand the vocabulary size to about 10000 words of each category resulting in a very high coverage of the Spanish telephone directory. The system uses a preselection module [2], using semi-continuous HMMs and a lexical access system. This module passes the 400 best candidates to a fine modeling module, which uses context-dependent semi-continuous (or continuous in a second version) word models. We will also include a spelling module in our system that will be used before falling back to the operator. We have laboratory results (not field tests) for this final system. Using the SpeechDat database we obtain a WER of 23.1% (first candidate) and 14.8% (two candidates). In the spelling module we have a WER of 12.7% (first candidate) considering the same vocabulary of 10,000 words.

4. CONCLUSIONS AND FUTURE WORK

Results from the first demonstrator show the feasibility of an automatic directory assistance service and a satisfactory user acceptance of such system even with sub-optimal components, as long as an operator fallback is provided for.

The common system architecture agreed upon proved successful. The modular design makes it possible to quickly build interactive telephone applications and to easily exchange modules (e.g. as was practiced between the Swiss and German partners).

Until the end of the current IDAS project, the second demonstrators with a substantially larger directory coverage will be completed and integrated into the real-life telecom environments where field tests will be carried out.

In a follow-up project real mixed-lingual speech input and output (e.g. for person and street names of non-native origin) will have to be incorporated as well as a more human-like dialogue modeling to allow a faster service and a higher automation rate at the same time.

5. REFERENCES

[1] IDAS-Interactive Directory Assistance Services Annual Report, 1999. European Project LE4-8315, DG XIII. Available from: http://www.HLTCentral.org/hlt/projects/idas/ar-99/ar-99.htm.

[2] J. Ferreiros, J. Macías-Guarasa, A. Gallardo, J. Colás, R. Cordoba, J.M. Pardo., and L. Villarrubia. Recent Work on Preselection Module for Flexible Large Vocabulary Speech Recognition System in Telephone Enviroment. In *ICSLP'98*, 1998.

[3] G. Hanrieder. Integration of a Mixed-Initiative Dialogue Manager into Commercial IVR platforms. In *Proceedings of IVTTA 98, Torino, Italy*, 1998.

[4] B. Kaspar, G. Fries, K. Schuhmacher, and A. Wirth. Spradiak - experience with the automation of the german directory assistance service. In *Proc. Workshop COST 249, Rhodes, Greece*, 1997.

[5] B. Kaspar, K. Schuhmacher, and S. Feldes. Barge-in revised. In *Proc. Eurospeech'97, Rhodes, Greece,*, 1997.

[6] J.M. Pardo, et al. Spanish text to speech: from prosody to acoustic. In *International Conference on Acoustic*, 1995.