

Experiences from a Danish Spoken Dialogue System

Hans Dybkjær
Prolog Development Center A/S (PDC)
H. J. Holst Vej 3C-5C
DK-2605 Brøndby
www.pdc.dk/speechlogic
+45 3636 0037
dybkjaer@pdc.dk

Laila Dybkjær
NISLab, University of Southern Denmark
Forskerparken 10
DK-5230 Odense M
www.nis.sdu.dk
+45 6550 3553
laila@nis.sdu.dk

ABSTRACT

Industrial speech recognition for the Danish language has been available for the last 2-3 years. We describe experiences from one of the first and so far most complex dialogue systems in Danish, aimed at providing general information about holiday allowance.

Categories and Subject Descriptors

H5.2 [Information interfaces and presentation]: User Interfaces - *Voice I/O*. D2.2 [Software engineering]: Design Tools and Techniques - *user interfaces*. H1.2 [Models and Principles]: User/Machine Systems - *Human factors*.

General Terms

Design, Experimentation, Human Factors, Languages.

Keywords

Spoken dialogue, industrial experience.

1. INTRODUCTION

For the last 2-3 years industrial level speech recognition for task specific natural language dialogue has been available for Danish. In year 2000 a market investigation of recognisers showed that there was one Danish recogniser, produced by Philips, for over-the-phone speech recognition. In autumn 2001 a Danish recogniser from Nuance appeared, also for over-the-phone speech, and now IBM plus NST seemingly are on their way. Philips also offers open microphone recognition, typically dictation systems specialised to a particular area.

We are aware of one such Danish dictation system which is meant for doctors. As regards spoken over-the-phone applications in Danish there are still only a few of them, including a country code information system, a flight timetable and arrival information system, an address information system, and the holiday allowance information system presented in this paper (+45 4820 4910 code 3).

A general interest in the area is emerging. But it is at the same time clear that there is quite some work to be done to cultivate a real market. The experience from other countries where spoken dialogue systems have been available for several years (the first

commercial system appeared in the US in 1989) is that such systems can certainly save money for the company and at the same time increase the service to customers. But there are many issues to be aware of when making a commercial dialogue application. Some of these dialogue application issues are discussed in books, such as [6]. We will address some of the issues we have experienced while developing one of the first and at the same time most advanced commercial Danish dialogue systems.

2. THE HOLIDAY ALLOWANCE PROJECT

The above-mentioned market investigation also looked at the market interest in spoken dialogue systems in Denmark. Information was collected via a questionnaire sent to about 200 companies and institutions. One of the very interested respondents was ATP (The Danish Labour Market Supplementary Pension Scheme). Together NISLab and ATP applied for money to initiate development of a spoken dialogue system. They obtained funding from The National Agency for Enterprise and Housing. After a call for tender the Danish software house PDC was chosen as software developer and the project started in autumn 2001. The agreed upon system was in the area of holiday allowance (Danish: feriepenge) which is administered by FerieKonto at ATP [3]. For all employees 12,5% of their salary is saved each month as holiday allowance. Employees continuing in a non-temporary position will just get their ordinary salary but if they change to another position or have a temporary position, they will get a holiday allowance certificate and will get allowance during their holiday by filling in this form and submitting it to FerieKonto. The rules concerning holiday allowance are quite complicated so many people have a need for information.

So far FerieKonto has via a voice-response system offered very general information on what to do if you have a holiday allowance form and on when you will get your allowance. In addition FerieKonto has offered access to a web page with answers to frequently asked questions.

ATP constantly aims at improving their service to customers while at the same time keeping the costs at a minimum or even reducing them. They also have as a goal to be on the front edge of development. Taken together this was the motivation for ATP for being so interested in a spoken dialogue project. They know that there is still a considerable number of their customers who either don't have access to the internet at home or who don't use the web pages even if they have access. On the other hand, nearly all people have a phone and know how to use it. However, making answers to frequently asked questions available over the phone requires much more than an ordinary voice-response system.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Second Danish HCI Research Symposium 2002, November 7, 2002, Copenhagen, Denmark.

Copyright 2002 Prolog Development Center A/S & NISLab/SDU.

3. TWO-STEP APPROACH

It was agreed to take a two-step approach to the holiday allowance system to be developed by PDC and NISLab. The first part of the system would be a fairly simple general information system which would only have a slightly larger coverage than the existing voice-response system for general information. We have called this first part Vejled (Guidance). The second step would enhance the first part into a Frequently Asked Questions system which we have named FAQ. The primary goals of Vejled were (i) to get the technology into place while still having a relatively simple dialogue system, and (ii) to generate initial experience with real users. We knew that the FAQ system would be a very challenging system to address with its unstructured task which is difficult to handle in a reasonable way without adding too many annoying dialogue constraints.

4. USERS AND DATA

The end-users for the system are FerieKonto customers, i.e. any person who has holiday allowance and has questions about the holiday allowance rules. The system is a walk-up-and-use system, which means that it is intended for users who have no prior knowledge of the system. Interaction is via speech only over the phone.

Another type of user is the system administrator at FerieKonto. His interface to the system is different from that of the end-users. It requires some training to learn how to administer the system, and interaction is via a normal GUI and not via speech. In the following we only consider the end-users.

We wanted to involve users from early on in the development process. As mentioned, in principle any employee is a potential user of the system. However, it is not without problems just to use any such user. The main problem is that they only have a fictive need for information expressed in the scenarios given to them in writing or orally. Also such users know they are part of a test and tend to be more patient than one would be in a real situation. For Vejled we started by using colleagues as test users. They were typically just briefly informed about the system and asked to call. We made a first few rapid iterations in this way. All calls were transcribed and analysed resulting in changes to the system. The amounts of data were small since people didn't phone more than once unless explicitly told to do so. During spring 2002 we invited people outside our sites to call the Vejled system. This resulted in 225 calls which were transcribed. Transactions in the dialogues were carefully analysed. By a transaction we understand a piece of information the user tries to get from the system. In one dialogue the user may make several transactions. A coding scheme was developed for the mark-up of transactions and a coding tool developed by PDC was used for the actual annotation. This coding tool includes statistics so that e.g. transaction success could be easily calculated. The transaction success was 91.8 for these calls.

People were encouraged to fill in a questionnaire but despite the 225 calls we only received 12 filled questionnaires. On the average users were quite positive but clearly there was still room for improvements.

We have also made closely monitored lab-tests. They require quite some effort and don't provide a lot of language data but per call they certainly provide by far the most information on which improvements to make.

Real data can only be collected with real users and that is what we are doing for Vejled now as reported in Section 5. For FAQ we are still running lab-tests and tests involving subjects who volunteer to call the system, but we expect to have collected more experience with this system before the symposium takes place.

5. OPERATION

At the time of writing Vejled has been in public operation for more than a month. Since Vejled has been put in operation we have collected almost 1000 calls to the system by the public which are pretty many calls taking into account that we are outside the holiday season. All calls have been transcribed and are now being analysed. We expect to report on the findings in these dialogues at the symposium.

6. TECHNOLOGY

Both Vejled and FAQ are implemented on the Philips SpeechMania platform [5], which includes a recogniser, language processing, and tools for output generation and transcription. In order to build a spoken dialogue system on this platform one has to implement the dialogue manager, add lexicons and grammars for the task in question and record output phrases unless one decides to use speech synthesis. Considering price and customer quality, recorded phrases were chosen for both Vejled and FAQ. However, the system is prepared for synthesis as parameters of synthesis continuously improve. In terms of development and maintenance and for dynamic systems synthesis is much to prefer.

The implementation is done in HDDL (higher dialogue definition language) which is an event-driven imperative language for dialogue management with a C-like syntax, supplied with declarative grammar and knowledge rules for modelling language and semantics.

The implementation employs standard hardware such as IBM Windows 2000 servers and NMS telephony boards.

7. DIALOGUE DESIGN

Two example dialogues are given in Table 1 at the end of this paper. Some of the design issues we have had to address include:

Introduction: The introduction must indicate to the users what they can expect from the system, and what they can say. Also, the introduction must be inviting. Many users hang up immediately during the introduction phrase, without trying to speak to system. The reader is encouraged to consider how the introduction in the Vejled example in Table 1 meets the requirements of being inviting, telling what can be done, how to do it, and not being too long, at the same time.

The "rhythm" of the dialogue: A natural dialogue must take into account issues such as the length of information blocks, when to inform and when to ask questions, detection of back-channelling, when to make pauses and how long silences to accept.

Help: First of all we need to detect that there is a problem. If a problem is detected, we need contextual instructions which offer more elaborate help than some general help function would do, and ultimately personal assistance should be offered.

Meta: In general, the ability to continue smoothly in case of miscommunication, and to make clarifications must be available. There are several examples in the FAQ dialogue below.

Language: Issues include vocabulary, syntax, and statistical evidence. The latter is very important in today's recognisers and requires large amounts of data. For speaker independent telephony applications transcription of 2-10 hours of user input from application realistic dialogues is needed. For dictation systems, 100-500 million words of text of the type to be dictated are required. The system should accept input as close to natural language as possible within the given domain, so transcribed material from user tests is also used to improve the system's coverage.

Feedback: The system must make clear to the user what it understood and what it is doing and why. The feedback may be explicit (query for confirmation), implicit (say what has been understood, but continue without waiting for confirmation) or implied (the very act of the system shows what has been understood). The FAQ example in Table 1 shows examples of all three cases.

Prompts: The system's vocabulary, formulations, voice and other effects such as music greatly influence both the user's perception of the system, the user's language, and the user's understanding.

Information management: The domain information, such as the holiday allowance rules, addresses, and opening hours, is quite huge and will even change over time. An issue is how to specify the dialogue so that it can be communicated to and understood by the customers and their domain experts, i.e. in our case ATP and people at FerieKonto, and so that it – once it is implemented - can be easily maintained. In the reported system we have defined a domain specific representation in XML, which is then compiled into the executable HDDL (see Section 6) code.

8. RESEARCH AND ENGINEERING

The basic recognition technology has been available for more than 10 years although considerable improvements have been made up through the nineties. Specifically for the Danish language the appearance of Danish recognisers within the last couple of years has been crucial to the step from research systems only to commercial applications.

In our experience the co-existence of research and engineering can be very fruitful. Often, research is ahead of technology. The basis of today's dialogue design was refined by research in the mid-nineties [1], but e.g. barge-in as well as the use of recognition scores for controlling feedback [2] have just recently become technologically reliable. Recent research demonstrates that prosody can be used for detecting corrections and aware sites [4], but today's recognisers do not support this feature. Here engineers owe a lot to basic research.

It also happens that technology is ahead of research. For example, input provides clues, such as references, negation and affection, but we do not know how to handle these in an operationally tractable and robust way. The problems have wider perspectives for our general knowledge, and the concrete problems observed in industrial applications may serve as motivation for research.

Finally, there is much more to making a customer application than what is addressed by research:

Telephony technology: Every new combination of PBX (switchboards), telephony boards, and IVR/speech platforms is almost certain to yield new, hard problems. Related is also dimensioning: How many lines, how many licenses, how many and powerful machines?

Real world size problems: The FAQ specification is some 50 pages long, and compiled into HDDL more than 15.000 lines of high-level code. This is a lot of data to explain to customers and co-developers and to administer during maintenance, and means of presentation and representation must be designed.

Testing: Does the application run robustly, meeting requirements of transaction success rates and runtime with no breakdowns?

Documentation: The system might have to be taken over by other people; legally it is not developed by single persons but by companies.

Administration and maintenance procedures: Both data handling, annotation tools and standards, installation and update procedures need to be specified and developed. Decisions must be made on work division between the software vendor and the customer's system administrators, and the latter must be trained.

Marketing and sales: Considerations of how to sell more are crucial.

Finalising: A research project can be a success even if it just ends giving some good experiences. An industrial project must end with a product that satisfies the customer.

9. ACKNOWLEDGEMENTS

The project is supported by The National Agency for Enterprise and Housing, The Danish Ministry of Economic and Business Affairs. We would also like to thank the staff at ATP for kind, competent and enthusiastic cooperation.

10. REFERENCES

- [1] Niels Ole Bernsen, Hans Dybkjær, Laila Dybkjær: *Designing interactive speech systems*, Springer 1998.
- [2] Hans Dybkjær, Niels Ole Bernsen, Laila Dybkjær: *Inferential issues in a spoken language dialogue system*, Proceedings of FQAS'96, Roskilde University, Denmark, 1996.
- [3] FerieKonto/ATP, www.FerieKonto.dk, 2002.
- [4] Julia Hirschberg, Marc Swerts and Diane Litman: *Labeling Corrections and Aware Sites in Spoken Dialogue Systems*, Proceedings of the Second SIGdial Workshop on Discourse and Dialogue, Aalborg, Denmark, 2001.
- [5] SpeechMania, www.speech.philips.com, 2002.
- [6] Susan Weinschenk and Dean T. Barker: *Designing Effective Speech Interfaces*, Wiley, 2000.

Table 1. Two example dialogues - translated from the Danish

<p>Vejled: Complete customer dialogue from the Vejled system.</p> <p>Very typically the customer calls, selects his/her situation, gets the information, and exits immediately. If the user does not hang up, the system will continue to offer operationally related information such as address and email-address, and to start all over again. Also not shown is the pop up of contextual instruction and help prompts whenever problems occur.</p>	<p>FAQ: Subset of user test dialogue from the FAQ system.</p> <p>This dialogue typically gets much more conversational, with topical changes and meta-communication. Note e.g. the extra confirmation in S 3.1, the prompting S 6.1 when the user says nothing, and U 7.1 where the user clarifies that s/he meant transfer of holiday. Also note the use of barge-in in U 7.1, a feature also present in Vejled.</p>
<p>S 0.1 - Welcome to FerieKonto's speech based guidance system</p> <p>S 0.2 - Here you can get standard information about payment of holiday allowance. - You may always ask for help or repetition.</p> <p>S 0.3 - Say if you - are an employee or unemployed in Denmark - are unregistered with the national register and live abroad - or have left the labour market due to age or health.</p> <p>U 0.1 I am unregistered with the nat- the national register <i>Recognised I am unregistered got the national register</i></p> <p>S 1.1 - Unregistered with the national register.</p> <p>S 1.2 - Do you live in Denmark?</p> <p>U 1.1 no <i>Recognised no</i></p> <p>S 2.1 - If you are unregistered with the Danish national register and do not work in Denmark</p> <p>S 2.2 - You can get your holiday allowance transferred to a Danish or foreign bank. - You must send your bank account number by e-mail, fax or mail. - Remember to write your civil registration number. - For foreign banks you must also write the detailed address of the bank. - A transfer to a foreign bank costs 75 DKK in exchange fee.</p> <p><i>Event 1 nothing recorded</i></p> <p>S 3.1 - If the FerieKonto certificate is correctly filled in when we receive it, we transfer the money to your account. - We need about five working days. - You cannot get your money until one month before the first day of your holiday.</p> <p><i>Event 2 hangup</i></p>	<p>???</p> <p>S 2.1 Ask about how you earn holiday allowance, get a FerieKonto certificate, fill in your certificate, send in your certificate, or get your money.</p> <p>U 2.1 how do I earn holiday allowance <i>Concept contribution(how do I earn holiday allowance)</i></p> <p>S 3.1 - You asked about - earning</p> <p>S 3.2 - Your employer must set aside 12,5% on top of your salary.</p> <p>S 4.1 - Some employers have an approved holiday allowance scheme and thus they need not pay to us. - In that case you must contact the employer directly to get a holiday allowance certificate.</p> <p>S 5.1 - You may also ask about problems with your employer's payment of holiday allowance or about transfer of holiday to the next holiday year.</p> <p><i>Event 4 nothing recorded</i></p> <p>S 6.1 - Sorry. - What did you say?</p> <p>U 6.1 I would like to know something about transfer of holiday allowance to the next holiday year <i>Concept accept(would like to)</i> <i>Concept next(next)</i> <i>Concept money(holiday allowance)</i> <i>Concept transfer(transfer)</i></p> <p>S 7.1 - The money is transferred to the account that appears from your FerieKon-</p> <p>U 7.1 @ transfer holiday <i>Concept reject(nope)</i> <i>Concept holiday(holiday)</i> <i>Concept transfer(transfer)</i></p> <p>S 8.1 - Usually you must take your holiday before the end of the holiday year. - However, your union may have an agreement that allows for transfer of holiday. - You may then arrange with your employer that holiday beyond 20 days is transferred to the next holiday year. - Otherwise we transfer your undrawn holiday allowance to the Labour Market's Holiday Fund which will use the money for collective holiday purposes.</p> <p>???</p>