

SPREIN - A VOICE I/O MAIL ORDER SYSTEM WITH TELEPHONE ACCESS

H.-W. Rühl*, L.M. Winzer*.

ABSTRACT

For a mail order application, the SPREIN voice I/O system is currently under development. It will accept orders via telephone without any human operator interaction. Potential users have to enter their orders by voice, and their speech will be interpreted by a speaker independent isolated word recogniser. In response, the SPREIN system will use a speech synthesiser capable to output unrestricted text on a phoneme basis.

The recogniser and the speech synthesiser are controlled by a dialogue controller interfacing users via a data line to the remote host computer of a mail order company. The remote host computer and the dialogue controller share responsibility for the ordering procedure. While the host computer manages the general flow of control for an ordering procedure, the dialogue controller is in charge of user guidance and assistance during user interaction.

INTRODUCTION

In 1985, the German PTT launched the SPREIN project to develop a voice I/O service system consisting of an unlimited vocabulary speech synthesizer, a speaker independent isolated words small vocabulary word recogniser, a dialogue controller, and a standardised interface to remote computers. After its completion, it will be installed at the PTT's local subscriber exchanges.

The objective of the system is its application as a user interface for a mail order shop computer during a one year's field test that is intended to start in the second half of 1987. But the long term interest of the PTT is to have a flexible voice I/O server system that may be offered as a service of the PTT and which may simply be adapted to the customers application as a front end.

SYSTEM STRUCTURE

The general structure of the system is depicted in fig. 1. It consists of a central controller handling the link to the remote host and the flow of control within a user dialogue in multiplexed mode, and one or more telephone conversation units TCU, each TCU forming the speech I/O interface for one telephone line.

Each telephone conversion unit consists of a word recogniser, a speech synthesiser, and a line interface, which are connected via separate serial RS232C lines to a corresponding dialogue controller task.

Speech output is generated by a SAMT phoneme synthesiser developed by the German PTT (ref 1, 2). With the help of the PTT's tools for gene-

* Philips Kommunikations Industrie AG, Kommunikationssysteme, Thurn- und-Taxis-Str. 14, P.O.Box 4943, D-8500 Nuernberg 10, West Germany

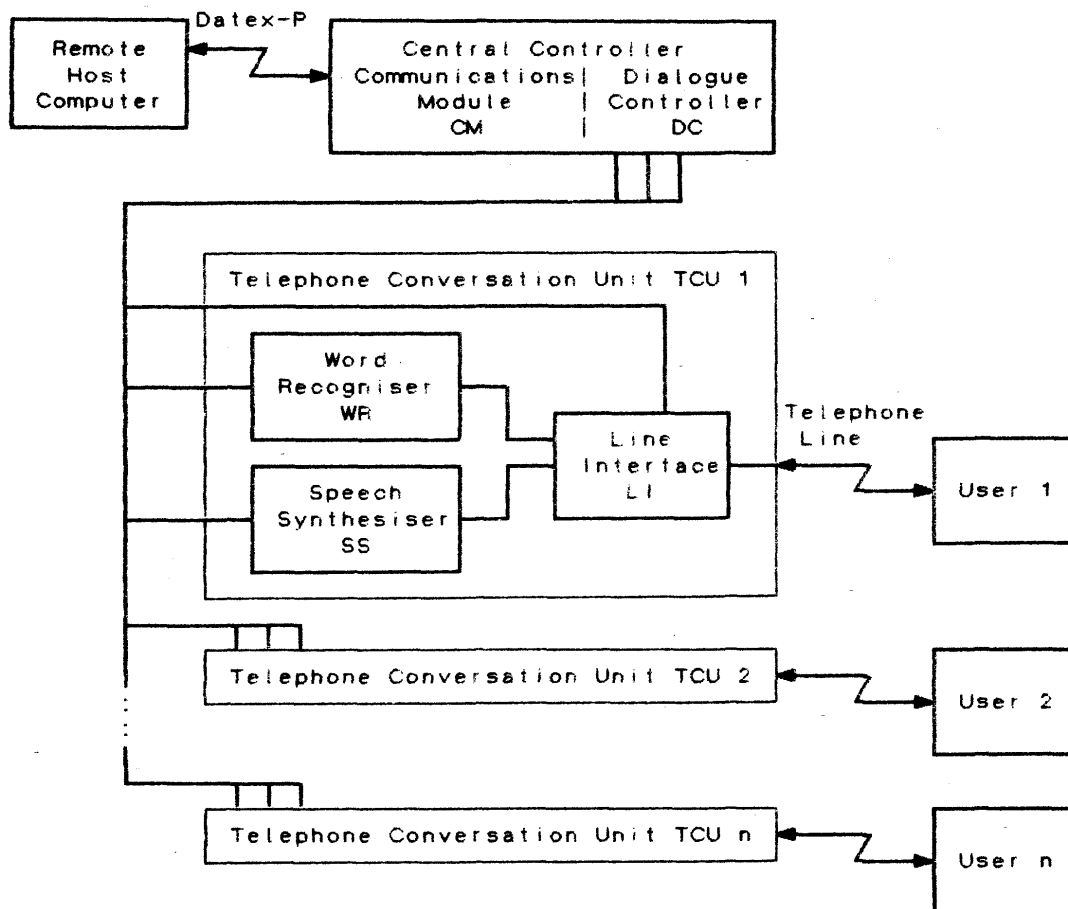


Fig. 1: General structure of the SPREIN voice I/O server unit

ration of phoneme strings including timing and intonation control from text, and for phoneme string optimisation, a large amount of messages may simply be generated or modified offline and then stored within the dialogue controller. To control the SAMT, an average data rate of about 200 bps is necessary. Hence, phoneme strings can easily be transmitted to the SAMT via a serial line in real-time.

The word recogniser is based on algorithms and software described in reference 3. It is a speaker independent isolated word recognition system designed for telephone line access. For the current application, a vocabulary consisting of the ten digits and seven other German words (Yes, No, Stop, Cancel, Pause, Repeat, Correction) with at most 14 active words is sufficient. Since the recogniser software runs on an MC 68020 μ P system supported by a TMS 32010 DSP system to compute a set of speech parameters every 12 ms, a gap of about 0.4 s between input words is needed solely for a correct end point detection even on noisy channels.

The central controller may run several orders consisting of a dialogue controller task and a communication task, one for each of the activated TCU's. Each communication task is connected via a multiplexed X.25 line (DATEX P) to a corresponding order task at the remote host compu-

ter. The X.25 line is run with an EHKP protocol (ref 4) designed originally for Bildschirmtext (similar to the British PRESTEL service) and extended for this application. Thus, one SPREIN system may handle several conversations. In the mail order application, only two TCUs will be used.

The central controller employs an MC 68010 μ P running under the operating system UNIX. Although UNIX is not a real-time system, its task switching time of few milliseconds does not cause any perceptible delay.

APPLICATION INDEPENDENT DIALOGUE CONCEPT

An important design rule for speech systems says: 'Tailor the recognition task around your application and use all the application specific knowledge obtainable to reduce recognition errors'. Because of this rule, most applications have dialogue controllers conscious of the contents of the items for which they are requesting. In an ordering task for example, redundancy in an article's number may be used to reduce the size of the input vocabulary, to inherently perform error checking or correction. SPREIN is different from this type of speech systems as it has been expected to be usable with different applications and hence may not have any knowledge about the context of the data that it receives.

This problem was solved, firstly by placing all of the general information into four application specific data files, and secondly by specifying a communications protocol between host computer and SPREIN capable to provide detailed format information via references to the elements of the data files. The data files are read once at the start of the system. So, to adapt the system to a new application, only the data files need to be changed and the system also needs to be restarted. As the data files are in textual form, they may be changed simply, allowing customers to prepare their own dialogue with no other tool but a text-editor.

The most important element in communication between host and SPREIN is an 'item prompt'. Item prompts are initiated by the host computer. They consist of an optional message field, a reference to a prompt unit, an input type, a respond vocabulary, and an input format specification. SPREIN then issues the optional messages, decodes the prompt unit to obtain user guidance information for the entry of the desired input data field, activates an appropriate vocabulary subset for each column of the input field. It finally sends the collected data back to the host computer, and awaits the next prompt demand.

The optional message field references announcements from a synthesiser message file holding at most 512 different messages accessible from the host only. It holds messages that cannot be related to a standard prompt unit such as "The ordered item is out of stock". A second file of same size exists which contains messages, that may be used only internally, i.e. via indices from a prompt unit as described later.

The response vocabulary index specifies the vocabulary subset allowed to appear in the collected data field sent from SPREIN to the host computer. Up to 256 different vocabulary subsets may be specified per

application. To determine an active vocabulary for the word recogniser, the dialogue controller adds words with only internal meaning (pause, repeat, correction etc.) to the selected response vocabulary.

An input format specifier is necessary to determine minimal and maximal length of the string to be input. With minimal and maximal length different, the user may input a termination command to finish after having entered at least the minimal number of words.

The prompt units stored in a prompt file contain references to all announcements necessary to support the entry of the expected input field. These are structured by recogniser control commands and flow of control commands. Prompt units may therefore be assigned meanings such as 'entry of account number'. Internally, they are divided in three sub units, i.e. entry of the first word of an input field, of the rest of the input field, and confirmation of the data field entry. The first subunit is mandatory while both of the others are optional. The confirmatory sub unit may be enabled or disabled on line by the host computer via the input type specifier.

A typical prompt unit in the mail order application will start by issuing a brief request for the item to be input. The request is given by a variable length string of indices to the internal messages file. It then makes the word recogniser wait for a user entry for the time specified by a recognition time command in the prompt unit. With no user response, additional information is announced and recognition is attempted several times as described in the prompt unit until finally either the user enters data or the system aborts due to time-out.

Special commands within a prompt unit may be used to issue messages for users whether familiar with the system or not. The last input word or the complete field as input so far may be echoed. Other commands control output of positional announcements ('please enter the *fifth* digit') or of announcements dependent on position ('you may terminate the string now by entering "stop"'), or with respect to a previous entry (confirmatory: 'you want to *continue/abort*').

CONCLUSIONS

No results are available yet, as the mail order field test is intended to start in the 2nd half of 1987 and will last for a year. Apart from performance data about the SPREIN I/O server, the test is expected to deliver general information about the public acceptance of voice I/O devices, and to gain experience in their use as a public service.

REFERENCES

1. H.E. Wolf: Entwurf und Realisierung eines Formantsynthesizers mit paralleler Filterstruktur für die Sprachsynthese nach Regeln. PhD thesis, TH Darmstadt, 1981
2. H.E. Wolf: Control of Prosodic Parameters for a Formant Synthesizer Based on Diphone Concatenation. Proc. ICASSP 81, pp. 106-109, 1981
3. M.H. Kuhn, H.H. Tomaschewski: Improvements in Isolated Word Recognition. IEEE Trans. ASSP-31 No. 1, pp. 157-167, 1983
4. Deutsche Bundespost: Bildschirmtext-Rechnerverbund Protokoll-Handbuch. Deutsche Bundespost, 1985.