

DELIVERABLE IDENTIFICATION

Identification number	LRE-63314-D1.1
Type	Technical Report
Title	Database Requirements for Teleservices
Status	Final Version
Deliverable	D1.1
Work Package	WP 1
Task	Task 1.1
Period covered	01 Jan. - 01 March 1995
Date	31 March 1995
Version	Final Version
Number of pages	6
Author(s)	Per Rosenbeck
Work package (WP) / Task (T) responsible	David Pearce (GEC) Per Rosenbeck (Tele Danmark, Jydsk Telefon)
Project contact point	Harald Höge, Siemens AG, ZFE T SN 5, D-81730 München Phone: +49 89 636 3374, Fax: +49 89 636 49802 E-mail: hoege@habicht.zfe.siemens.de
CEC project officer	José Soler
Status	Public
Actual distribution	Consortium and CEC
Supplementary notes	
Key words	
Abstract	
Status of abstract	

Received on	
Recipients catalogue number	

DOCUMENT EVOLUTION

Version	Date	Status	Notes

Database Requirements for Teleservices

0. Identification

Task Manager:	Tele Danmark, Jydsk Telefon
Task Partners:	Philips, Alcatel, Vocalis, INESC, IDIAP, Siemens, GEC, CSELT
Task:	1.1
Start:	1. January 1995
End:	1. March 1995
Effort:	1.7 MM

1. Introduction

A prerequisite for a successful automated teleservice is robust speech recognition. However, robust speech recognition for teleservices heavily relies on the availability of suitable speech databases, which are required to train the speech recognizer for a specific application. The objectives of this report is twofold:

- To analyze the market demands in the range of teleservices and their impact on the design of the speech database.
- To specify the impact from current and near-future speech recognition technology on the design of the speech database.

In the first objective, a selected number of current succesful teleservices world-wide will be described with respect to the knowledge we have on the service itself and the associated database. These succesful services also points out directions of the market demands, and a prediction of near-future teleservices will be done in this chapter.

The second objective reflects the new possibilities and challenges that arises from new advanced speech recognition technology, and its influence on the design of speech databases. This chapter will give an overview of this technology, and come up with some examples of impact on the database design.

The requirements outlined in this task will contribute to the specification of the database (Task 1.3) and the case study for the creation of short-term SLR (Task 1.4).

2. Analysis of current teleservices

This chapter will analyse the scope and market of a selected number of teleservices, major services will be identified and their impact on the database design will be clarified. The topics that is addressed in the conclusion of this chapter is:

- Definition of vocabulary for teleservices

- Required coverage of accents and regional languages
- Profile of expected users

Since the creation of a database is a major effort, it must be carefully analyzed which automated services the database is intended to support. Examples of the following teleservices will be considered with respect to a clarification of the database requirements and the market demands:

- services for public switches, eg. directory assistance and voice dialling
- services for private switches, eg. automatic call distribution
- remote access to a private database, eg. voice mail and information retrieval

The customer needs and market demands in this range of teleservices have to be taken into account in the design of the teleservice and the associated speech database. The following section describes a selection of some successful teleservices.

2.1 Description of major succesful teleservices

This section describes a selection of some succesful teleservices seen world-wide. Each of these teleservices is described in a table, stating the name of the service, the manufacturer, the contact person and a short description of the application and references. This is followed by estimated figures for the number of current users, and the succesful transaction rate. The table then describes the user dialogue and finally a description of the associated database, with respect to the number of speakers, regional language areas, vocabulary size and content, distribution og age and sex, and the method of recording. The following list of applications is selected for this description:

- StockTalk, Bell Northern Research, Canada
- VRCP, AT&T Bell Lab, USA
- Traintable Info System - Philips, Germany
- Telia - Directory Assistance service, Sweden
- Deutsche Bundespost Postbank Telebanking System, Germany

Name of teleservice	StockTalk
Manufacturer	Bell Northern Research (BNR), Canada
Contact person	Matthew Lennig, BNR
Teleservice description	Used for Stock Quotations in Canada and selected areas in USA. New company names are included every week. Quotations is obtained by calling a telephone number and pronouncing the company name. Listing is done by a Text-to-speech synthesizer. Have been an actual running service since '92. User investigation conducted '94.
References	Presented at ICSLP 92, Banff, Canada
Number of users	Telephone users in Canada and USA
Successful transaction rate	89-90 %
Recognition technique	Flexible vocabulary, based on CHMM phoneme recognition
Dialog description	Dialogue presents StockTalk, and prompts for company name.
Database description: - Number of speakers - Regional languages - Vocabulary size/content - Distribution age/sex - Method of recording	A total of 127 speakers (half male and female) were used for creating CHMM phoneme models. Collected in the Montreal area. Each speaker read his/hers 40 specific company names over dialed-up lines. A variety of telephone sets and office environment were used. Total of 1561 different company names were recorded. Training set consisted of 2212 tokens from 59 speakers.
Estimated Revenue	

Name of teleservice	Voice Recognition and Call Processing (VRCP)
Manufacturer	AT&T, Bell Lab
Contact person	Jay Wilpon, Bell Lab
Teleservice description	Processing of calls by use of ASR- eg. Third party call, credit card call etc. Possible to get connected to an operator - but a large part of the calls are automatic
References	ICSLP-94 in Japan
Number of users	Several states in the US - increasing rapidly
Successful transaction rate	> 95%
Recognition technique	Whole word and garbage modelling
Dialog description	Simple introduction and wait for user reaction
Database description: - Number of speakers - Regional languages - Vocabulary size - Vocabulary content - Distribution age and sex - Method of recording	A lot of the speech material are collected "on the fly", ie while the service is operational. The collected speech material is thereby very realistic and gives a high performance and robustness rate.
Estimated Revenue	

Name of teleservice	Train timetable information system
Manufacturer	Philips Research Laboratories Aachen
Contact person	Dr. Martin Oerder
Teleservice description	Automatic inquiry of train timetable information of the German railway using a natural language interface.
References	e.g. Paper on last ICSLP
Number of users	Up to now, a few thousand have tried out the system
Successful transaction rate	about 75%
Recognition technique	large-vocabulary continuous-speech recognition
Dialog description	natural, no rigid dialogue structure
Database description - Number of speakers - Regional languages - Vocabulary size - Vocabulary content - Distribution age and sex - Method of recording	bootstrap db creation: a new version of the system is trained by the speech material recorded from users so far.

Name of teleservice	Telia - Directory Assistance
Manufacturer	Vocalis/Logica
Contact person	
Teleservice description	Automated directory assistance Automated welcome greeting and number release, user voice requests for return to operator or repetition of information - Language: Swedish. Date of service: April-November 1994 - Public field trial for subscribers in Stockholm, Uppsala and Vasteras January 1995 - Public nationwide service
References	"Speech Technology: A One Shot Possibility". Bo Forssten, Telia TeleRespons AB. Computer-Telephony Conference, Voice 94 Proceedings. 1994.
Number of users:	
Successful transaction rate	Recognition accuracy 97%
Recognition technique	CHMM phoneme-based word spotting recognition with talkover and rapid vocabulary building
Dialog description:	
Estimated Revenue	Projected operational savings of US \$6 million 10% increase in efficiency
Database description: - Number of speakers - Regional languages - Vocabulary size - Vocabulary content - Distribution age and sex - Method of recording	

Name of teleservice	Deutsche Bundespost Postbank Telebanking System
Manufacturer	Controlware/Syntellect
Contact person	Peter Keppler, Training Manager, Controlware, Germany
Teleservice description	Voice Recognition Telebanking Service. Language: German. Date of service: January 1994 - Public nationwide service - Balance request - Statement Order - Demo-system - Marketing information request about services - History of telephone accesses - PIN recognition for secure access and for PIN alteration - 24 hours service except "Transfer to Agents"
References	"German Postbank 'Telebanking' System". Peter Keppler, Controlware. Computer-Telephony Conference, Voice 94 Proceedings. 1994.
Number of users	650,000 customers in July 1994 (16% of customer base), 600,000 requests for balance information per month
Successful transaction rate	
Recognition technique	Isolated word recognition of German digits 'Null' to 'Neun' and 'Ja' and 'Nein'
Dialog description	
Database description: - Number of speakers - Regional languages - Vocabulary size - Vocabulary content - Distribution age and sex - Method of recording	

2.2 Database requirements

The above described successful teleservices opposes some requirements on the contents of the associated databases:

- Phonetically rich sentences, in order to design teleservices with phoneme based CHMM, and building of rapid vocabularies
- Command and common used control words for the teleservices
- Word spotting material, ie command and control words in context, ie. embedded in typical sentence structures and from spontaneous and realistic user reactions

3. Speech technology impact on database design

The second dimension along which a database has to be specified, are the technology constraints. The database has to be designed with respect to the potentials and limitations of current and near-future speech recognition technology. Whereas current systems often rely on isolated word recognition in the context of a rigid dialog, future systems will be more user-friendly. Clearly, there is a need for telephone databases with spontaneously and realistic spoken speech phrases. Also the issue of achieving a phonetically balanced database or at least a phonetically rich one, is important in the design of robust teleservices, including a subunit-based recognizer. The most popular subunits used is phonemes, either in context (biphones or triphones) or context-independent ones. Some of the trends seen from the major successful services described in chapter 2 is that:

- Continuous-speech recognition (no pause between words required) will become prevalent
- Commands/keywords can be embedded in carrier phrases
- The user will be able to barge in, i.e. to interrupt the announcement instead of waiting for the system prompt
- Market demands on robust and reliable teleservices
- The user will be able to speak in a natural, spontaneous manner and provide/ask for information in any order, not being restricted to a prespecified dialog structure.

The speech technology that goes along with these trends are the following:

- Continuous Speech Recognition technique (coarticulation handling, subunit modelling)
- Word spotting technique (embedded keywords/commands in phrases)
- Barge-in technique (interruption of announcement in user-machine dialogue)
- Methods for achieving robustness and noise reduction techniques, in order to achieve a higher successful transaction rate for a given teleservice
- Design techniques for a more natural (efficient) user-driven dialogue instead of the traditional machine-driven dialogue, but to keep in mind - simplicity.

In many cases additional material can be collected “on the fly”, while the teleservice is operational. This additional material will be more “realistic”, since it is recorded in optimal environment, namely the actual teleservice (-or at least a prototype of the service). With the additional material you can improve your speech models and in this way the overall performance and robustness of the teleservice. In the US it is legal to do this extra recording, if you notify the users that they are being recorded. But to use this approach you will have to record a speech database that can help you launch the teleservice.