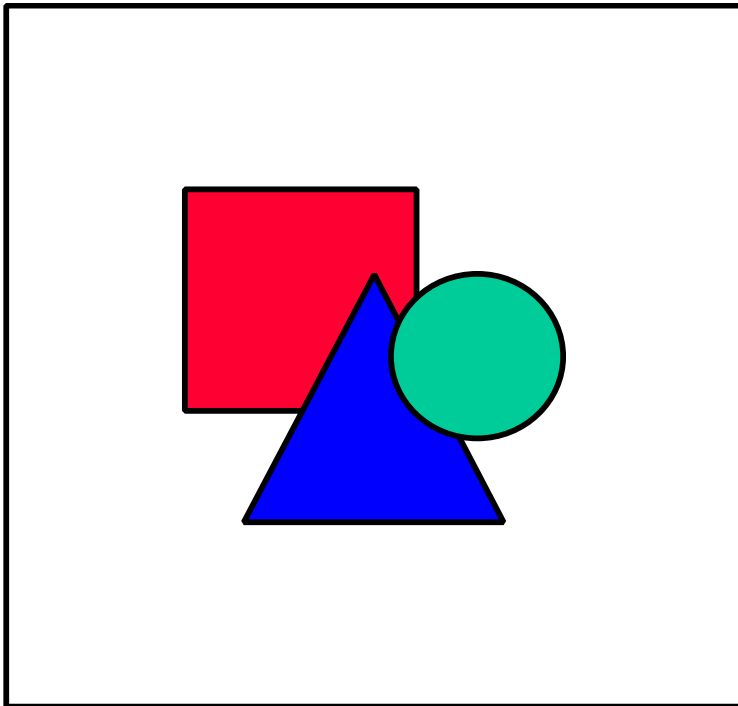


LE1-1930

FINAL REPORT

PUBLIC VERSION



**TABLE OF CONTENTS:**

1	EXECUTIVE SUMMARY	2
1.1	OBJECTIVES AND ACHIEVEMENTS	2
1.2	VALIDATION RESULTS.....	2
1.3	IMPACT AND FUTURE PROSPECTS	3
1.4	PROJECT TIMETABLE	4
1.5	STAGES OF WORK	4
1.6	EXTERNAL REVIEWS, CONFERENCES, EXHIBITIONS, USER GROUP MEETINGS	5
2	ACHIEVEMENTS.....	5
2.1	MAIN FUNCTIONS SUPPORTED.....	5
2.2	TECHNOLOGIES AND COMPONENTS	6
2.3	SOFTWARE AND HARDWARE REQUIREMENTS	7
2.3.1	<i>Architecture review.....</i>	<i>8</i>
2.3.2	<i>Basic Speech Server System Capabilities</i>	<i>9</i>
2.3.3	<i>Standards.....</i>	<i>10</i>
2.4	PREREQUISITE SOFTWARE AND PORTABILITY	10
2.5	PROJECT RESULTS.....	11
2.5.1	<i>User Requirements.....</i>	<i>11</i>
2.5.2	<i>Technology Improvement.....</i>	<i>11</i>
2.5.3	<i>Human Factors.....</i>	<i>12</i>
2.6	OTHER RESULTS	12
2.7	PROJECT-LEVEL DISSEMINATION, AWARENESS, PUBLICATIONS, ETC.	12
2.7.1	<i>Contribution to standards.....</i>	<i>12</i>
2.7.2	<i>Awareness.....</i>	<i>13</i>
3	EVALUATION AND ASSESSMENT	13
3.1	VALIDATION AND FEED BACK	13
3.2	INTERNAL COLLABORATION:	14
4	CONCLUSIONS AND FUTURE PROSPECTS	14
4.1	TECHNICAL FEASIBILITY, IN TERMS OF MATURITY, SCALABILITY, OF THE DEMONSTRATOR	14
4.2	ECONOMICAL VIABILITY, BASED ON A COST/ BENEFIT APPRAISAL	15
4.3	BUSINESS PERSPECTIVES	15
4.4	ANTICIPATED DEVELOPMENT OF THE MARKET SECTOR(S) ADDRESSED	15
4.5	RECOGNITION, IMPORTANCE AND SUPPORT ATTACHED TO THE PROJECT, IN-HOUSE; INTEREST SHOWN BY PROSPECTIVE CUSTOMERS.....	16
4.6	EXPLOITATION PLANNING	16
5	APPENDIX.....	2
5.1	TECHNICAL DESCRIPTION	2
5.1.1	<i>Introduction.....</i>	<i>2</i>
5.1.2	<i>Hardware Overview.....</i>	<i>3</i>
5.1.3	<i>Hardware components</i>	<i>3</i>
5.1.4	<i>Software Description</i>	<i>5</i>
5.2	SPEECH RECOGNITION RUNTIME SYSTEM	6

This work has been supported by:

- the Telematics programme of the European Union
- the Office Federal de l'Education et de la Science (in Switzerland)



1 Executive Summary

1.1 Objectives and achievements

The work in CAVE has substantially contributed to:

- obtain a better understanding of user requirements and marketing issues related to SV as one means of protecting tele-business applications from fraud
- develop operational SV technology that can actually be deployed
- perform tests of SV technology integrated in tele-services

The experience gained in building the services will be extremely valuable for the development of speech driven services by the service operators.

The SV technology developed performs an a par with -if not better than- the best technology available elsewhere. The research platform developed in CAVE that allows the partners to share work in the development of SV technology has proven to be extremely effective, and has aroused substantial interest in the research community.

The technology is undoubtedly ready for deployment in well designed –but low risk- applications.

Most important in the definition of SV- facilitated access procedures to tele- services is:

- Ease of use and robustness
- Single enrolment procedure
- Service integration including appropriate fall-back provisions
- Well defined management procedures (e.g. real-life database management)

The demonstrators have been built in client-server architecture. The SpeechServer was hosting the speech recognition and speaker verification resources. A separate platform was hosting the telephony boards (clients) and the application. The application has been developed with commercially available service creation tools.

1.2 Validation results

The objective of CAVE was to specify and implement two pilot versions of telecommunication services deploying speaker verification (SV) technology:

- one in the field of telephone banking
- and one using a Calling Card service.



The focus of the CAVE project was not to develop new applications based on the SV technology, but to find out customer reactions and attitudes towards this new technology.

Two validation set ups have been implemented.

- The telephone banking application was run in Switzerland.

Users got fictive bank accounts and their voices were used to identify the owner and the account, as well as to do speaker verification. After users were verified properly, they could transfer money between their accounts. The language used in the banking application was Swiss German.

- The calling card application service in Holland;

Users were speaking Dutch and entered the calling card number with their voices. The calling card number was evaluated with speech recognition and the users were verified with speaker verification on the given number.

The reactions of users are very influenced by the performance level of the technology and by the quality of the design of the human-machine interface of a service. User reactions and user-classification in different categories (goats or sheep) depending on their ability to cooperate with the system was also by using information about objective system performance collected in the two field tests.

1.3 Impact and future prospects

The speaker verification technology as has been developed under CAVE might already be appropriate in a number of low risk tele-services. More elaborated verification strategies however are required for advanced tele-businesses. Seamless integrated speech recognition and speaker verification technologies might facilitate extremely friendly but well secured accesses to information and transaction services. Enrolment and robustness of SR-SV combinations should be appropriate and improved within applications.

Reduction of fraud and advanced user-friendliness are the main drivers to applications in:

- New services. To improve image and to generate new profits.
- Existing services. To achieve more ease of use.

<i>Telecom</i> <i>Banking</i> <i>Electronic Commerce</i> <i>Penitentiary</i> <i>Applications</i> <i>Physical Access</i> <i>Etc.</i>



1.4 Project Timetable

With hindsight, we can say that the planning has proved to be too tight after all. However the original course of the project was sound, and it has proven successful. Unexpected problems in delivering and tuning the test systems have really affected the overall progress of the project. After ample discussions it has been decided to remain on the initial tracks. No major reorientation has been felt necessary.

Start of the project: 01-12-95
End of the project 30-11-97

1.5 Stages of work

The preparations of work have been started already in 1995 resulting in a 'hot start' early 1996. Initial activities regarding market research and user requirements research have been started simultaneously with preparatory work on the functional specification and technical specification of the demonstrators to be built.

- First demonstrators (laboratory versions) became available in the second half of 1996. The mimicked versions and field test versions came available not before mid 1997. Significant problems related to speech recognition and system integration had to be solved.
- Final user tests have been performed per end 1997. Final evaluation including the preparation of all associated deliverables was in the first months of 1998.

The demonstrators in Switzerland and in the Netherlands will be kept operational during the first stages of the successor project PICASSO.



1.6 External reviews, conferences, exhibitions, user group meetings

The work of the CAVE team has been 'followed' and 'evaluated' by the responsible officers of the European Union and according normal practice by external reviewers also.

Two user group meetings have been organized.

- the first mainly to arise awareness in partner organizations;
- the second mainly to achieve interaction with external institutions and potential customers. Also first introductions on privacy issues (legal aspects) in speaker verification applications have been held.

The meetings and reviews have been of great value for the CAVE consortium.

CAVE has successfully contributed to a number of national and international events.

Apart from scientific goals much effort was spent also on the creation of user awareness and to verify the assumptions made from within the CAVE project.

A selection:

- | | |
|---|-----------------------|
| <input type="checkbox"/> Security and Comfort | Brussels 1995 |
| <input type="checkbox"/> Computer Security Ass. | Utrecht 1996 |
| <input type="checkbox"/> AVIOS | San Jose 1996 |
| <input type="checkbox"/> AVBPA97 | Crans Montana 1997 |
| <input type="checkbox"/> ICASSP'97 | Munich 1997 |
| <input type="checkbox"/> EU Banking Day | Brussels 1997 |
| <input type="checkbox"/> Eurospeech '97 | Rhodes 1997 |
| <input type="checkbox"/> EU TAP98 | Barcelona 1998 (spec) |

WEB presence <http://www.ptt-telecom.nl/CAVE> was given from the early start of the CAVE project.

- Overviews and outlooks are presented through the video produced at the end of the CAVE life cycle.
- Demonstrators have shown very important to show the potentials of SV and to achieve fruitful discussions with service providers.
- On a national scale all partners have put significantly effort in presentations and demonstrations.

2 Achievements

2.1 Main functions supported

The assumption at the start of the CAVE project was that speaker verification is a viable means for reducing fraud in telematics services, even if it had not been widely tested in close-to-real applications, for reasons that were not well understood. Thus, the major aims of the project were:

- To obtain a better understanding of user requirements and marketing issues



- To develop operational SV technology that can actually be deployed
- To perform tests of SV technology integrated in tele-services

Thus, CAVE has addressed user requirements analysis, the development of demonstrators as well as the improvement of existing technology, and the validation of the concepts relating to suitable use of the technology and the technology itself in small scale user tests.

2.2 Technologies and components

Speaker Verification technology comes in three flavors: text-dependent, text-prompted and text-independent. The test applications developed in CAVE call for the use of text-dependent and text-prompted approaches.

Customer's voice models can take several different forms. In CAVE an approach was further developed that allows to formulate virtually all promising approaches in terms of (variants of) Hidden Markov Models.

To normalize for a multitude of fundamentally uncontrollable conditions the similarity of a speech token to a speaker model is estimated as a ratio between the speaker-model likelihood and a world-model likelihood.

In developing the technology two corpora have been instrumental, viz. the American YOHO corpus, and the SESP-1 corpus previously collected by PTT Telecom. Furthermore, the research aimed at improving the performance of existing SV technology was based on a common software platform, which is based on the Hidden Markov Model Toolkit (HTK) by Entropic.

In the interface to a telematics transaction service SV almost by necessity must be combined with some form of automatic speech recognition (ASR). In CAVE an existing HMM-based ASR engine has been used. Based on that engine digit and yes/no recognition has been developed for Swiss German and Dutch.

Even though the same HMM technology is used for speech recognition and for speaker verification, the technologies differ.

- the ASR module must derive the textual representation of a speech utterance,
- in the SV module the text of the utterance is known a priori,

The CAVE R&D system has been built to allow either cohort models and world models, but ultimately we decided to focus on the tuning and improvement of the world-models approach only.

Main arguments for this decision are:

- A world model is much less time consuming in terms of computation, since only one likelihood value has to be computed for the anti-speaker likelihood



- A world model is much more economical in storage volume than a client dependent cohort model
- A world model does not require a selection of cohort speakers (and no well-established procedure for doing so exists)
- The performance of a speaker independent world model has never been proved to be worse than even the most elaborate speaker dependent cohort systems

In an application using SV three discrete phases must be distinguished:

- Enrolment
- Threshold setting
- Access

All three phases have been addressed in CAVE. The major results are summarized in next paragraphs.

2.3 Software and hardware requirements

In CAVE a client - server approach has been implemented for demonstrator building. The servers offer speech recognition and speaker verification functionality to any machine networked to them via the Dialogic SCx-bus for guaranteed real-time transmission of digitized speech data). This architecture is suitable for commercial deployment, as it is stable and scaleable.

Each server offers its resources for some number of independent "channels", thus allowing for multiple simultaneous connections to the server. The clients to this server are telephony-equipped PCs, each running an application that answers the telephone to incoming callers and maintains the dialogue with the user, securing and using a speech server resource only when necessary.

The scalability of the Dialogic SCx-bus allows a simple test system based on a single client and a single server to grow to meet additional demands.

In CAVE speaker verification algorithm development and improvement has used a standard laboratory infrastructure for speech technology research, viz. UNIX workstations. The research platform used the HTK toolkit from Entropic Research Inc.

For a number of reasons CAVE has deliberately chosen to build any of its demonstration systems on top of existing Vocalis technology.

The application programming interfaces (API) -already existing at Vocalis for application development- are rather quickly extended to provide all the functionality required for the building of the CAVE test systems.



2.3.1 Architecture review

The Speech Server architecture enables efficient sharing of expensive speech recognition (SR) and speaker verification (SV) resources among multiple Voice Response Units (or Telephony Servers).

Based on the client/server model, the Speech Server architecture distributes the basic voice processing tasks among specialized systems to achieve better overall performance and resource utilization.

This architecture can also be extended to enable one or more servers to act as Intelligent Peripherals in telephone network applications.

Speech Servers are also ideally suited to speech technologies prototyping and evaluation because they offer a well defined, distinct and confined, controlled environment to experiment with innovative hardware and software.

In a typical Speech Server configuration, Telephony Servers are client systems that hold telephony hardware and run the application itself.

They are networked to Speech Servers, server systems equipped with speech recognition and/or SV hardware and software

There can be a pool of clients and only one server, or a farm of servers and one client, or even a pool of clients and a farm of servers.

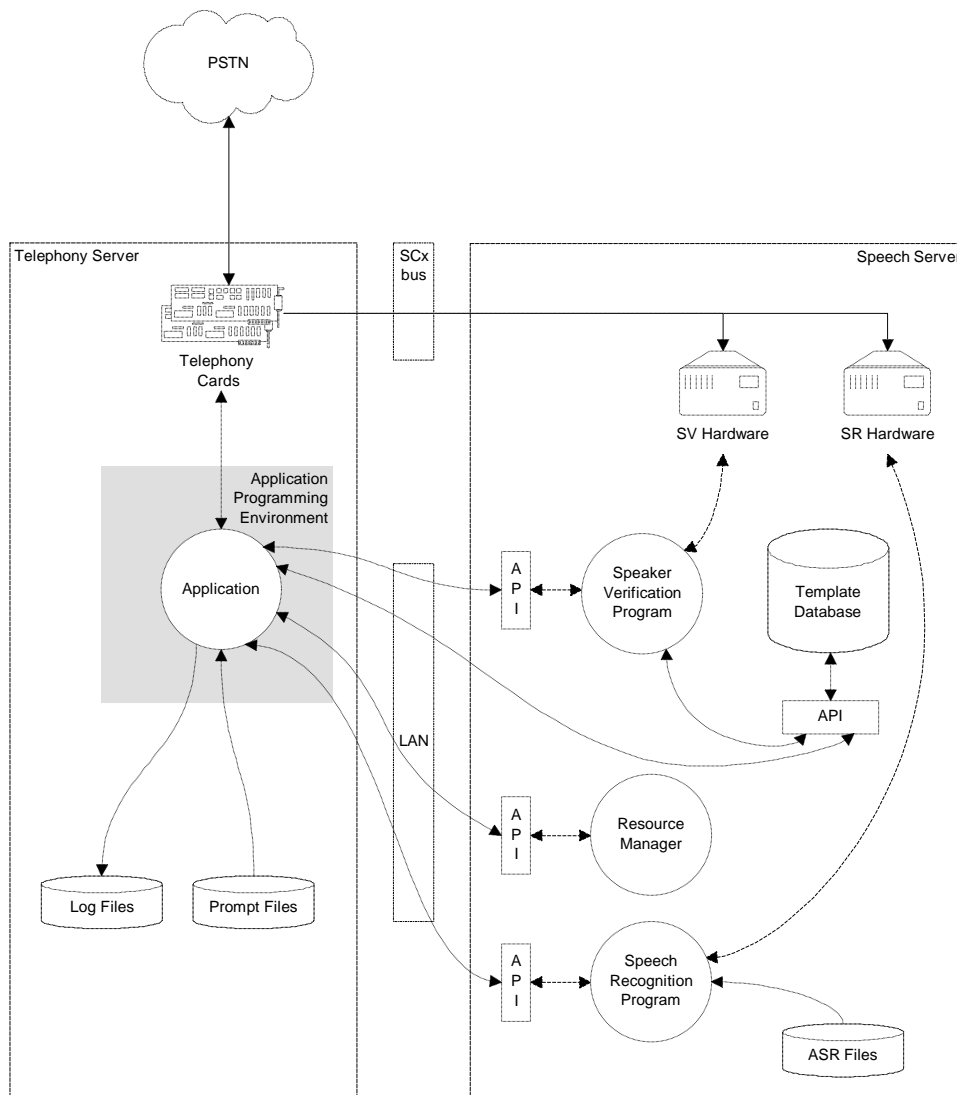
Such details are totally transparent to the application. Simple configurations with one Telephony Server linked to a single Speech Server are also supported.

Clients don't necessarily have to be networked together, but need to be linked to every single server using a TCP/IP-UDP/IP LAN or WAN.

Although data and commands are transmitted through the LAN, the recognizer gets its speech data for recognition purposes from an external speech bus: the SCx-bus.

As a result, clients and servers are doubly linked: by a data-only TCP/IP-UDP/IP network and also by an external SCx-bus speech bus.

Speech Servers can in principle be integrated in infrastructures with existing non-Vocalis speech recognition resources and coexist with them. Applications can even be written that interleave recognitions on legacy recognizers with recognition on the Speech Server.



2.3.2 Basic Speech Server System Capabilities

The Speech Server capabilities are outlined below:

- support for speech recognition
- support for text-dependent SV
- support for text-independent SV (not required by CAVE)
- support for text-prompted SV (not required by CAVE)

There were some constraints that limited the range of options available to design the CAVE demonstrator. These were required for reasons of more rapidly achieving the project goals and were:

- real-time operation
- use of existing recognition software
- non-existent SVAPI standards
- limited development time and resources



The initial systems were based on existing technology as far as possible to ensure robust, real-time operation, with resources devoted to integration of the SV components and basic system support functions. Once this was realised, a platform was available for development of the more sophisticated features of SV and ASR ultimately recognised as being needed to meet the project goals.

2.3.3 Standards

The Speech Server architecture is based on open industry standards:

- UNIX operating system,
- TCP/IP-UDP/IP protocols,
- resilient industrial PC platform,
- Ethernet data networking,
- SCx-bus speech bus (Dialogic SCSA hardware),
- C language software.

This ensures that it meets the requirements for a large class of industrial applications.

SCSA (ECTF) had been specified in the Technical Annex as the intended standard to which CAVE project developments should adhere if possible.

At present the SCx-bus and hardware standards have been well-defined and are adopted by a number of CTI vendors and manufacturers, including Dialogic who provide a wide range of industry standard SCSA-compatible boards suitable for project use and approved for installation in many countries.

2.4 *Prerequisite software and portability*

Except for the widely available (almost 'academic standard') HTK package, the CAVE research platform does not require any software that is not available on a standard UNIX workstation. Much effort has been spent to ensure that the CAVE reference systems produces identical results, irrespective of the hardware platform on which it is run.

The HTK scripts developed in CAVE will be made available to the research community, to foster R&TD in the field of speaker verification and to promote the use of formal evaluation methods. Moreover, the SESP corpus on which much of the technology development in CAVE was based will be made available to the research community via ELRA.



2.5 Project results

2.5.1 User Requirements

The single most important results of the user requirements research was the finding that user friendliness is as much a concern in secure transaction services as safety. Service providers consider SV as a means for simplifying the user interface, rather than as an essential safety protection measure. Existing services already have some form of protection, based on other techniques or procedures, and SV might be deployed to get around some of the less user friendly alternative protection measures. Of course, this may be different for completely new services.

Another finding was that --at least for the time being-- fraud is not generally considered as very urgent. Of course, every provider of transaction services would like to reduce fraud to the very minimum, but the profit margins of the existing services do already take some amount of fraud into account. Until competition forces providers of financial services to reduce their costs quite strongly, the motivation to invest in additional protective technology for existing services will be limited. However, here again, the introduction of a fast growing set of home banking services is likely to spur widespread use of user-friendly biometric protection techniques, first and foremost SV.

The user requirements research has also pointed out several new and creative ways of deploying SV, integrated in a service in such a way that it need no longer be used as a binary accept/reject device. The alternative use triggers an alarm if the verification confidence is below a threshold, leaving it to the service or the service provider to take appropriate action.

2.5.2 Technology Improvement

Research in CAVE has resulted in very substantial improvements of the base line performance of the SV technology that we had available at the beginning of the project. Equal Error Rates have been reduced with an order of magnitude, first on the YOHO corpus (clean speech, and therefore not representative of what is to be expected in telematics transaction services), subsequently on SESP, a realistic corpus of telephone speech. One way in which the performance has been improved is by inventing better ways of estimating the co-variance matrices in the speaker models, despite the lack of sufficient enrolment speech to obtain reliable estimates directly. A patent application covering these inventions has been filed.

In addition to off-line optimization of the technology, our we have obtained a much better understanding of the impact and implications of a priori estimation of accept/reject thresholds in the absence of sufficient enrolment data. Presently, the feasibility of a patent application relating to the threshold estimation techniques is under consideration.



2.5.3 Human Factors

Experiments with the two demonstrator systems have brought to light the strong dependence of an interface using SV on the performance of an ASR device, that is needed to produce a completely speech driven service. Mixing speech and DTMF is very confusing for most users, and should therefore be avoided.

One experiment has shown that users lose confidence in SV at the first time they are confronted with a false accept; false rejects are very unpleasant, but at least in our experiments they were not considered as prohibitive. However, we expect that the latter will not hold in real services.

Subjects accept two short or one relatively long (> 4 minutes) enrolment session(s).

Unfortunately, even two sessions is not enough to cover even the larger part of the between session variation in handsets, channels, acoustic backgrounds and speaker status that are likely to be seen in an actual service.

2.6 Other results

An analysis of the use of calling cards has shown that many customers use their card to call only a very small number of connections. In the calling card demonstrator this finding has been used to reduce the number of recognition errors in telephone numbers: subjects were asked to submit the list of numbers they expected to call, and these were integrated into the language model of the recognizer. A patent application has been filed covering the creative aspects of this procedure to overcome the limitations of the present ASR technology in a calling card service.

In a commercial version of the service number dialing might be replaced by name dialing.

The work on improving SV technology proper has resulted in a very general and easy to use tool that runs on top of the HTK software package that supports R&D in speaker verification. This tool, as well as the procedures developed in CAVE to reduce error rates on pre-recorded corpora will be made available to the research community. The same holds for the SESP-2 and SESP-3 SV corpora that have been collected in CAVE.

2.7 Project-level dissemination, awareness, publications, etc.

2.7.1 Contribution to standards

CAVE has made some contributions to the definition of a standard API for speaker verification.



2.7.2 Awareness

In addition to the formal User Group meetings CAVE (and the individual persons working in the project) has made many contributions to the public awareness of SV technology and its potential applications in telematics transaction services and electronic commerce.

Collaborators in the CAVE project have given numerous interviews to journalists writing for the public and the popular scientific press. Last but not least, a video presentation has been produced, that has already been used to promote SV technology.

3 Evaluation and assessment

3.1 *Validation and feed back*

The research and the applications build have been validated in a number of sequential tests. A generic SV Research Environment has been developed and distributed to all involved Universities and Research Institutions.

Reference databases (Yoho and SESP) and well-defined methodologies and communication procedures have created a most efficient atmosphere for simultaneous validation of the research.

In each of the two business areas (banking and telecommunications) two systems have been built.

Those systems have been tested in three different environments:

- A laboratory environment (to evaluate the technology of a demonstrator)
- A mimicked environment (to evaluate the system integration)
- A field test environment (for user tests)

The dependency of two new technologies (speech recognition and speaker verification) in the CAVE applications influenced the possibilities to get applications ready in time to perform tests, collect material, making interview and draw the right conclusions as was planned. Most of the work however have been completed before the end of the CAVE life cycle. The final reports on evaluation have been completed.

In the two fieldtest several verification and enrollment strategies have been experimented.

The banking application run in Zurich. In total about 185 'speakers' have enrolled into the system. The calling card application run in The Hague. Some 100 users have been enrolled into the system.

All enrollments and accesses to the services have been analyzed by analyzing the log files made and by collecting questionnaires. The results are reported in a number of deliverables.

All calls have been recorded and will be analyzed for the PICASSO project.



3.2 Internal collaboration:

CAVE has demonstrated the immense benefit that can be gained from an atmosphere of open collaboration and communication in EU funded LE projects.

To achieve the common goals CAVE partners (users, research and technology) have shown a permanent determination for sharing information, experience, success and (sometimes) failures - without reluctance or mistrust.

4 Conclusions and future prospects

4.1 Technical feasibility, in terms of maturity, scalability, of the demonstrator

Several issues emerged as a result of this work, which require special attention, and may be expected to be the subject of future research and development.

1. The provision of closely integrated, accurate speech recognition is essential for successful deployment of speaker verification applications. Commercial deployment of SV require more open systems. A modular design of speech functions and well defined technical interfaces are to be prioritized.
2. Recognition approaches and accuracy intrinsically affect speaker verification and impostor algorithms. This includes the detailed procedures for incorporation of silence, world and filler models in the scoring algorithm,
3. Rapid recognition response is needed for good dialogue control, but has to be balanced against accuracy of recognition, especially when lengthy verification utterances are used which may have longer inter-word pause durations
4. Thresholding procedures and algorithms are very important and require continued detailed investigation
5. The SV performance is very dependent upon appropriate quantities and types of training material. This has implications for service provision in terms of human factors and user expectations. This also is an important area for continued future research.
6. The API may require to be redeveloped in future work in the light of emerging ECTF standards.
7. The demonstrators implemented a relatively simple database management system based on the Unix system. Extension toward realistically sized applications will require significant engineering of the database functions for maintaining customer records, billing control, speaker templates, speech data etc. This is a major task requiring careful specification, and is critically linked to the required number of users/lines, choice of operating systems and available DBMS products.
8. Suitable management procedures are to be developed by the service providers



4.2 Economical viability, based on a cost/ benefit appraisal

Speaker Verification can be used in several application domains.
Main objectives might very divers; from fraud protection too ease of use.
The economical viability is to be analyzed per application.
CAVE has made some first exercises that showed its positive value.

4.3 Business perspectives

The CAVE project has generated knowledge and a number of technical achievements. These technical achievements include SV software, databases and the development of a generic SV research environment.

Knowledge is focused mainly on proper understanding of SV algorithms and associated improvement and human factors. The improved technology has been implemented and tested in a number of different verification strategies and applications.

This implementation includes integration in the demonstrators and applications built but also the embedding of SV in the existing SS1 architecture.

Partners in CAVE might have following business perspectives

- Consultancy on technological aspects
- Consultancy on SV related business and application development
- Benchmark testing/evaluation methodology
- Provisioning of SpeechServer equipment with embedded SV
- Provisioning of sophisticated SV software (without embedding)
- Reference databases
- Generic SV research environment

CAVE also gained huge amounts of knowledge on human factors research, among which dialogue design. In the project a unique cooperation between partners of very different nature has been achieved. Of course, the business perspectives differ between the partners, because of their different nature (technology providers, service providers, research institutes). Yet, the prospects are very promising.

4.4 Anticipated development of the market sector(s) addressed

The SV technology is still in its infancy; the same holds for our understanding of when and where the technology is used most profitably. Low risk applications seem to be the first applications that will go on-line. In medium to high risk applications SV can be used in combination of other (sometimes-existing) security measures. Embedding SV in existing human-machine interfaces might really lower thresholds to the effective use of tele-information and -transaction services.

A better understanding of business processes related to identification and verification will also support effective use.



CAVE did focus on applications in the business sectors of telecommunication and banking. Most important future technological issues seem to include:

- Incremental enrolment
- Better understanding of access/ verification strategies
- Robustness of ASR and SV
- Application Programming Interfaces
- Factors related to scaling to real-life technical and managerial environments
- Database management and customer procedures

The CAVE project was focused on the business area of banking and telecommunication. The incorporating of SV in generic e-commerce and web based procedures is the ultimate goal of CAVE and its successor project PICASSO.

4.5 Recognition, importance and support attached to the project, in-house; interest shown by prospective customers

In CAVE the Research and Academic partners have achieved the best SV performance rate ever been published. Through the commercial partners in CAVE awareness has been created for commercial deployment. In-house projects are under planning. This should be intensified in the PICASSO project.

4.6 Exploitation planning

Most service operators in the CAVE consortium are involved in tele-information and tele-transaction services and consequently they are interested in all potential measures to reduce risks and to create more user-friendliness in the human-machine interface. Image building and potentially the development of new business concepts by means of SV are also main reasons for embracing the opportunities of SV soon.



5 APPENDIX

5.1 Technical Description

5.1.1 Introduction

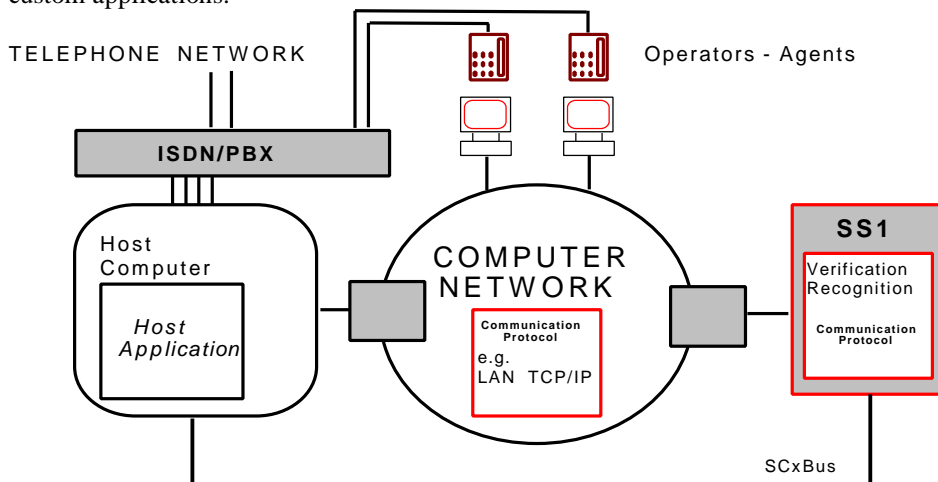
The Vocalis CAVE Speech Server (SS1) is an interactive voice response platform which delivers a range of speech technologies primarily to support speaker verification pilot and demonstration systems. It has been specifically designed to support real-time application development, experimentation and evaluation of speaker verification technology within the CAVE project. The SS1 incorporates Vocalis proprietary speaker independent speech recognition and CAVE speaker verification algorithms, as well as providing some general functions such as speech storage. (In principle it could also provide a full application run-time and development environment, supporting DTMF input, audio playback and telephony functions as a standalone IVR system, but this has not been required).

The speech server platform is designed to be used either singly or in multiple server configurations to provide scalability to large systems, and will interface with a number of different application host environments. The host machines will primarily implement all the dialogue and telephony functions of the application, including all supporting administration functions, leaving the server to provide only the speech verification and recognition functions. This architecture permits one basic server to be integrated with a wide range of applications and host environments.

The delivery platform is principally designed as a relatively low-cost unit suitable for small installations or research prototyping, and is not supplied with the full set of reliability features or documentation normally required for an operational network service. However the software and hardware configuration can be delivered in a fully specified Vocalis network platform should this be required.

Based on Vocalis's proprietary speech technology software, the SS1 can service a range of interactive spoken dialogue applications. It is intended that software upgrades will be produced during the lifetime of the project to enable continual developments in speech technology algorithms to be readily implemented, although the basic hardware should remain unchanged, with the exception of items such as the SCxBus card, planned to provide integrated digital telephony communications as this is developed later in the project. Access to Vocalis proprietary algorithms and library functions will be provided only through provided software interfaces. Source code will not be supplied other than as developed in the CAVE project.

The SS1's hardware and architecture as well as its UNIX operating system software are all industry standard, which maximises open-interconnectivity. The SS1 software is supplied with a library of modules, functions and tools implementing the CAVE Speech Server API functions, thereby allowing rapid development of custom applications.





5.1.2 Hardware Overview

The SS1 hardware is modular, enabling a wide variety of systems to be configured. The hardware is based on a standard tower or similar chassis with both Industry Standard Architecture (ISA) bus and PCI slots, with some room for expansion if required.

Basic computing resources are provided by a Pentium processor with 64MB random access memory, DAT tape cartridge drive, IDE floppy disk controller, SCSI hard disk controller, network card, SVGA display card and an SVGA monitor and keyboard. The SCSI controller supports both the hard disk and tape cartridge, and permits use of additional disk drives and CD-ROM reader for installation. A 2 Gb hard disk will be supplied to support the system services, storage of speaker recognition and verification models, and associated call speech data. This is expected to be more than sufficient storage to support all phases of the CAVE field trials, including recording of all speech data. A modem facilitates installation, support, updates and reconfiguration remotely by Vocalis.

Additional disk drives, ISA peripherals and interface cards can be connected as required by individual applications – depending on availability of back-plane slots and software driver support. Telephony interfacing and speech input/output on the host is to be managed by industry standard digital telephony cards. Provision is made for later addition of an SCxBus digital interface card, but initial system specification does not include this card, and TCP/IP connectivity will be used for both data and speech communications.

5.1.3 Hardware components

Item description	Qty	Reference**
Chassis	1	Standard tower or similar chassis, ISA and PCI slots, 3.5" 1.44 Mb floppy drive, with country-specific connectors
Single board computer	1	Pentium 120 Mhz ¹ , 512 Kb L2 cache, floppy controller, serial, parallel and PS/2 mouse ports
RAM	4	16 ² Mb (72 pin SIMM 9 chip 70 ns parity)
Hard disk controller	1	DPT Smartcache IV PM2144W PCI SCSI
Hard disk	1	2 Gb SCSI
Tape backup	1	DAT 4 Gb, internal
Network adapter	1	SMC 8216 ISA Ethernet 10 Mbit/s
Video controller card	1	SVGA
Modem *	1	External US Robotics Sportster 28.8 v34 (UK approved only)
Monitor	1	14" color SVGA
Keyboard	1	with standard connector
Mouse	1	Microsoft with PS/2 connector
SCxBus card *	1	Dialogic SCxBus adapter - can be added later as required

* Optional features not included in the standard specification and price quotation.

** Manufacturers and item types may be replaced by suitable alternatives

¹ This CPU is specified for a hardware-based recognition system. CAVE systems were actually supplied with a host software-based recogniser which required a 200Mhz Pentium Pro.

² 64 MB RAM was supplied for the host-based recognition system. The number of active recognisers is dependent upon RAM size.

Caller Verification in Banking & Telecommunications

Project : LE1-1930
Deliverable : Final report
Reference : D_FINAL



The following table describes the SS1's physical characteristics and the range of environmental conditions within which it can operate :

Temperature	0 - 55 °C. Reliable disk reading/writing 5 - 45 °C.
Humidity	0- 95% at 40°C (non condensing).
Altitude	15,000 feet / 4,500 metres.
Power Supply	95-132/180-246 V AC, 47-63Hz, 250 Watts
Physical **	254mm H x 2254mm W x 406.4mm D. Additionally, a keyboard and monitor.
MTBF	Manufacturers' figures for continuous 24 hour operation: Central processing Unit 50,000 hours Power Supply Unit 100,000 hours
Access **	0.5 metre envelope to sides, top, front, with 50cm at rear. SS1 is relatively portable, so access can be achieved by small movement of the unit, though no cables may be disconnected.

** Specifications subject to change if chassis is replaced by a suitable alternative

Speech Recognition

The SS1 is configured with a software speech recognition and verification system which may be shared across all four telephony channels, subject to reasonable resource loading. Hardware speech recognisers are not supplied as standard in the SS1, though they may be added subsequently, as needs dictate. These can be quoted for separately.

In most applications a recognition process is required for a relatively small proportion of the dialogue with the remaining time being spent in prompt payout. In this way the single software recogniser can generally service all 4 telephone channels. The number of simultaneous recognition processes possible without degrading response times is determined by the vocabulary size and recognition mode (i.e. isolated or continuous word recognition and word- or phoneme-based vocabulary).

The SS1 is delivered configured for recognition vocabularies up to 100 words Extending the vocabulary size in a software recogniser tends to introduce constraints in other parts of the system, such as response time or number of telephone channels which can be serviced.

The following speech recognition capabilities are supported:

Speaker Independent Recognition

- Flexible vocabulary creation based on phonemes (either British English, Dutch, Swedish or high German. Other languages can be supplied in consultation with Vocalis)
- Isolated word recognition - digits and keywords/keyphrases
- Keywords - yes, no, help, stop, cancel, repeat. A number of keywords can be supported using phoneme models for recognition if whole word models are unavailable.
- Continuous digit recognition.
- Top scoring candidate hypotheses with scores
- Dynamically loaded syntax trees
- Multiple mixture models

Single (later multiple) front-end parameterisation for recognition and verification



The following capabilities will also be provided in later releases of the software:

- N-best recognition, with resulting scores
- Improved word and phoneme models in formal German for Swiss subject to database provision by CAVE partners
- Improved CAVE algorithms and additional front end parameterisations

It may be possible to provide the following capabilities if required, in a separate quotation:

- Word spotting of words or phrases surrounded by any other words.
- Out of vocabulary rejection
- Talkover

Speaker verification

- Enrolment and authentication modules for speaker verification using CAVE algorithms
- Single Gaussian mixture models
- Single (later multiple) front-end parameterisation for recognition and verification

5.1.4 Software Description

Item description	Qty	Reference
Operating system	1	SCO OpenServer 5 Enterprise - 2 users licence (UNIX)
Telephony drivers	1	Dialogic System Release 4.2 for SCO UNIX
Recognition (runtime)	1	Vocalis Speech Recognition Runtime System (includes Speaker Verification)
Recognition (development) *	1	Vocalis Speech Recognition Development System.
Template database*	1	Informix C-ISAM 3.0 or C-ISAM 6.0

* Optional features not included in the standard specification and price quotation.

The operating system is SCO Openserver 5 (UNIX). This affords connectivity to a multitude of third party communication products which interface to LANs, WANs and other communication interfaces.

Below the operating system layer a number of device drivers handle the telephony interfaces, recognition processors and host communications hardware.

Above the operating system, the API layer provides a library of C subroutines for speech recognition, verification and basic host communications. This API will conform to the CAVE specification, initially developed to a design specified solely by project partners. Later releases will be provided implementing the final CAVE API design.

Also included is Management Information Support offering features such as call statistics collection and line and system management functions. These features are set out below:

System Administration Interface

The SS1 provides a system administration interface which allows a system administrator to:

- Perform the system shutdown (this will allow the system to be powered down.)
- Change the system administrator's password.
- Archive log files and allow an operator to extract log files from the system.
- Carry out any application-specific system administration functions.



Call Logging

Each call creates a call record entry which is kept in an ASCII file on the SS1. It is possible to process call records to provide customised reports for detailed analysis. The call record will be defined by the CAVE project and contains test and debug logging information such as:

- The date of the call.
- The time of the call.
- The total duration of the call
- The server functions called and their activity status
- The incoming line number or telephony channel on which the call was taken.

5.2 Speech Recognition Runtime System

The Runtime system provides the following:

- utilities to configure and load the speech recognisers installed in Speech Server with vocabularies to be recognised
- a utility to monitor the usage of the speech recognisers in Speech Server
- Speech recognition firmware
- a watchdog process to monitor recognisers in Speech Server and reboot them if they are found to be down

That Runtime system also includes full support for SV.