



**LE1-1930**

**Initial Functional Specification of Demonstrators**

**Deliverable D 3.1 public version**



---

**This is the public version of the deliverables D 3.1A and D 3.1B (confidential)**

---

**TABLE OF CONTENTS**

1. General Overview .....	3
2. Description of Services .....	3
3. Functional Specification .....	5
3.1. User interface .....	5
3.2. Registration / Enrolment .....	5
3.3. Verification .....	7
3.3.1. Initial calling card set-up .....	7
3.3.2. Initial banking set-up .....	8
4. Auditing .....	9
4.1. Verification performance .....	9
4.2. System-user-interaction .....	10
4.3. Other functionality needed for evaluation .....	10
5. System Performance .....	11
5.1. Overview .....	11
5.2. Error rate .....	11
5.2.1. Speech recognition .....	11
5.2.2. Speaker verification .....	11
5.2.3. Response Time .....	11
6. System Integration .....	12

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. General Overview

The task of WP3 is to develop functional and technical specifications for the CAVE speaker verification systems in a banking application and in a telecommunications application. It aims to define the Technical Specifications of the sequential versions and implementations.

The CAVE project will implement demonstrators in three stages, in order to:

- provide prototype systems for early feedback into design and service provision
- incorporate more advanced SV algorithms as they are developed in WP4 during the project life-cycle
- incorporate system improvements (functional and technical, including human factors) as a result of experience gained with initial demonstrators
- support later planning and design of systems for in-service integration

The three demonstrator stages are nominally:

1. standalone demonstrator of SV functionality
  - basic SV enrolment and authentication functions
  - not necessarily including speech recognition
  - not necessarily implementing target application functionality
  - not integrated with existing service
2. standalone demonstrator of basic application incorporating SV
  - including speech recognition if required by application
  - implementing (mimicking) main target application functionality: adaptation, rejection during the enrolment
  - not integrated with existing service
3. fully integrated demonstrator system
  - implementing full target application functionality
  - integrated with existing service

This document describes the important issues associated with first laboratory versions of demonstrator to be build for banking domains and telecommunication domains.

The structure of this document is:

- description of service
- functional specification
  - user interface
    1. registration/ enrolment
    2. verification
- auditing
- system performance
- system integration

Due to the many simultaneous actions in the CAVE project it is to be expected that specifications might change rather often.

## 2. Description of Services

---

## Caller Verification in Banking & Telecommunications

**Project** : LE1-1930  
**Deliverable** : Initial Functional Specifications (first phase)  
**Reference** : D 3.1 - public version

---



This document specifies the functionality for a speaker verification-based access service to automated 24-hour banking and calling card information and transaction services.

It does not specify however the precise nature of the actual service or the information offered.

Instead, the emphasis is on speaker verification as a means to help prevent fraudulent entry into a system where sensitive customer data, or the ability to perform significant transactions, is on offer.

As well as speaker verification, the system should optionally offer speech recognition for service navigation where the use of DTMF (dual-tone multi-frequency, or touch-tone) digits is not appropriate or available.

The specification of the access and the enrolment modules are presented, including the details of the speaker verification and optional speech recognition. Access and enrolment will be made available in several languages.

All CAVE partners will be involved in development and building and testing of the several banking and telecommunication demonstrators. Actual building will be executed in close co-operation between the technology partner and the user partners in the CAVE consortium.

In order to determine and improve the quality of speaker verification first demonstration models will be built and tested in a laboratory environment.

The objectives of the first demonstrator models are:

- The development of a real-time system to demonstrate speaker verification
- The investigation of the system, hardware and software issues associated with speaker verification
- The demonstration of the capability of the system to business units
- The running of a field trial to evaluate speaker verification performance, user interface issues, and customer reaction
- The identification of the capabilities and limitations of speaker verification in real applications

Assuming positive performances sequential demonstrator models will be built and evaluated in real world environment of actual banking and telecommunication services.



### **3. Functional Specification**

#### **3.1. User interface**

It is envisaged that speaker verification will be differently integrated in banking and telecommunication services.

In the banking environment SV will be used as an initial security check in conjunction with existing methods and procedures. The banking service is intended to be fully automatic with no possible intervention from a agent throughout the call.

In the telecommunication environment SV will be used as an additional security concept complementary to existing procedures. Intervention and guidance from human operators / agents should be possible as usual.

'Integration or de-coupling' of SV technology and remainder of applications is highly affected by these factors. In the document we will not focus on the exact form of applications to which users will gain access through speaker verification. Description of actual applications has been avoided and / or limited to those details only necessary for better understanding of the specific SV approach.

Hypothetical- real-world-systems may use an automatically-controlled dialogue between the caller and the service, which allows to obtain :

- large amounts of caller's speech during arrangement of transactions
- additional user information

With first demonstrators we aim to investigate and demonstrate the appropriate balance of technological and procedural choices which are certainly application and service dependent.

This section describes the functional specifications of the service enrolment and access functions as well as the auditing which must be performed to record details of every verification attempt.

#### **3.2. Registration / Enrolment**

New users have to be registered by the SV system during a repeated so-called enrolment session. Enrolment includes training of the system and generating of representative voice profiles of individual card/ account holders. After enrolment the card/ account holders can use the access-service without undergoing this procedure again.

The number of enrolment sessions depends principally on the number of distinct telephones from which the automatic service will be accessed.

The content of an enrolment is dedicated and depends from the design of a typical application.

In contrast with a calling-card application it is fair to assume that for the purposes of the banking demonstrator the majority of calls from a single caller will come from a single work telephone or some small number of domestic telephones. Two enrolment sessions also arguably represents the best trade-off between verification performance and the complexity of the entire enrolment procedure for end users. (Although in a real service it may be possible to offer differing enrolment strategies appealing to varying levels of technical literacy in the customer base, the marketing issues associated with this should not be underestimated).

In calling card applications the variance in telephone sets and transmission lines could be much higher. This will be reflected in the number of repeated enrolment sessions and probably also in the complexity of the content of the enrolment.

---

## Caller Verification in Banking & Telecommunications

**Project** : LE1-1930  
**Deliverable** : Initial Functional Specifications (first phase)  
**Reference** : D 3.1 - public version

---



Regardless of the number of sessions, enrolment will involve the training of the system so that a voice profile, based on the card / account holder's utterance of the digits from zero to nine, can be generated. In addition, the card / account holder will be required to give a password that will also be profiled in this way. This password can theoretically be any utterance of sufficient length, such as a digit sequence, phrase or even sentence.

In the initial version of the demonstrator, speaker verification performance is to be tested in different verification scenario's. (text-independent or text-dependent or prompted)  
The explicit collection of password and digit utterances or profiles will allow for testing different scenario's using password/ phrase or digit utterances or combinations such as a spoken PIN code, or some number of randomly-chosen digits from it). The digit strings elicited from the caller will be ordered to cater for intonation effects.

An initial security key or protected procedure is necessary before actual enrolment starts.  
Before each individual enrolment session, a caller's identity must be verified through some other means.

Rather than using the card / account number and some 4-6 digit PIN code for this purpose, callers should be authenticated using some specially generated "session key". This would only need to be known for enrolment sessions and could be safely discarded when enrolment was finished. This key could be generated from card / account number and other personal data and entered on the telephone keypad using DTMF.  
It will be assumed for the purposes of the demonstrator development that all telephones will be able to generate DTMF digits.



### 3.3. Verification

Referring to the objectives of the first demonstrators a flexible dialogue structure for the communication between the caller and the system has been designed.

Depending from the results in workpackage 4 and 7 more explicitly recommendations or decisions concerning information flow during verification could be taken. Application dependent recommendations are to be expected.

#### 3.3.1. Initial calling card set-up

In the initial calling card set-up following structure will be valid:

Note: to isolate questions arising from SV and SR evaluators will be provided with an unique DTMF code representing their unique cardnumber and pincode. This of course during first laboratory tests only.

The systems checks the identity of the user with a DTMF code,

- The system asks for the cardnumber.
  - *The caller says his 14 digit cardnumber.*
    - Digit recognition and a Luhn-check will be used in order to determine if the cardnumber is valid.
    - If the cardnumber is valid, speaker verification will be used to determine if the caller is in fact the card holder.
      - If positive the system will thank the caller for participation and will terminate the call.
      - If not-positive the system will invite the caller for the pin-code.
  - *The caller says the personal pin-code.*
    - Digit recognition will be used in order to determine the pin-code.
    - If the pin-code is valid, speaker verification will be used to determine if the caller is in fact the card holder.
      - If positive the system will thank the caller for participation and will terminate the call.
      - If not-positive the system will invite the caller to repeat/ say a prompted sequence of digits
  - *The caller says / repeats the prompted sequence of digits.*
    - Digit recognition will be used to determine and verify the repeated sequence.
    - If the sequence of digits is OK, speaker verification will be used to determine if the caller is in fact the card holder.
      - If positive the system will thank the caller for participation and terminates the call.
      - If not-positive the system will repeat once the total dialogue.
      - If not-positive the second time the system will thank the caller for participation and terminates the call.



### 3.3.2. Initial banking set-up

In the initial SV secured banking set-up following structure will be valid:

Notes: = In the initial version SV will be used as additional security measure only.

- = 'Traditional' DTMF account numbers (bank account) and DTMF pin-numbers (service type) will provide generic access rights
- = SV will be implemented 'around' spoken passwords supplementing the account/pin number sequence.
- = Later versions might integrate spoken passwords and account/pin number sequences.

A typical flow is as follows:

- The system asks for the accountnumber.
  - *The caller keys his 9 digit accountnumber.*
  - The system will prompt for the pin/ service code.
    - *The caller keys his 4/6 digit pincode*
      - If the combination of accountnumber and pin code is not valid the dialogue will be repeated.
      - If the combination of accountnumber and pin-code is valid the system will ask for the spoken password.
    - *The caller says his personal password*
      - Speaker verification will be used to check the spoken password
        - If positive the system will thank the caller and gain access to the actual service (or in the demonstrator terminate the call)
        - If not-positive the system will ask for repeated input or terminate the call if verification is not successful two times.

In general terms errors are to be processed as follows:

1. An invalid account number or PIN number will lead to the account number prompt being played again, up to two times before the call is terminated.
2. If nothing is entered, whether by DTMF or voice, at any point, the system will prompt again for the input up to two times and then terminate the call.
3. Five seconds will be allowed for the user to start entering DTMF digits and three seconds between each digit. These values will be configurable.
4. If the speech input for the password is too quiet or too loud, the caller will be asked to repeat the password up to two times before the call ends.
5. If the caller is authenticated on the strength of the account number/PIN code combination, but speaker verification fails, the call will be terminated



## 4. Auditing

To be able to evaluate the real world performance of the system, a range of data from the user interactions with the system has to be logged and later evaluated.

The main types of logging data needed refer to two main groups

- verification
- usability

Both are related to different parts of the evaluation in WP7 of the CAVE-project.

These two logging-groups should be combined in the same outputfile. Dedicated datasets could be achieved by appropriate filtering.

A professional methodology is to be adopted for measurements of the behaviour, performance and improvements of SV technology in real life applications.

In next phases the CAVE consortium will pay attention to specifically those requirements.

In this document the requirements of laboratory phases has been emphasized.

### 4.1. Verification performance

Verification-performance logging information will be used to evaluate the performance of the verification in a real world environment. Experience indicate that the performance in real life will be considerable lower than in laboratory tests. It is therefore important to get knowledge about the performance in a more real world environment.

For each call, the logged information should consist of at least the following:

- System decision
- Accept/Reject threshold on the verification score
- The verification score
- All the speech uttered during the call
- Timing information
- Claimed user identity
- System version.

System decision, threshold and score information are necessary to determine speaker verification performance during the trial, and so that off-line calculations can be made to determine what effect changes in the rejection threshold would have on this performance.

Speech data will be needed in order to evaluate false acceptance rates. To make these evaluations, specific user models are put "on trial" with speech from other users, in order to simulate deliberate break-in. A sensible way to do this is to record all the field-test speech, thereby creating a speaker verification database. This data must be recorded in such a way that all relevant call log, timing, and user identity data can subsequently be associated with it.



## **4.2. System-user-interaction**

The system-user-interaction logging data is needed for evaluation of the usability of the system. It will be used in addition to user questionnaires in the evaluation in performed WP7 of the CAVE project. The logging data needed is the following:

1. Interaction technique - DTMF or speech, if chosen by the user
2. Path taken by the user in the service with times for different dialogue steps

Consequently the information to be logged should be:

- The decision made by the system
- The threshold used
- The score value that the threshold operates on
- Speech-data recorded during the use of the system, failures included
- Absolute times for start/stop of interaction
- The recognised cardnumber and pin-code
- Version of the system
- Caller-ID

## **4.3. Other functionality needed for evaluation**

Having full and detailed logging data is one of the prerequisites before evaluation can actually start. But large amounts of data implies that good filtering of logging data has to be possible when it is time for evaluation. Examples of filtering are:

Concatenate subparts of dialogue times. All information about a certain user, a certain dialogue part or transaction performed during a certain period. Statistics of the verification results must be possible to filter out with various constrains.

Functionality has to be present in the system to by force altering the users possibility to get into the system. This is due to the need to do test of the users reactions when being denied access to the system. That is denying the user access one or several times during a certain call, regardless of the verification results. This could be achieved by pre-programming that certain calls, in chronological order, denies access to the user, regardless of the verification result. The actual sequence of failing verification has to be decided by test specification in WP7.

It may be desirable to deploy certain features which will allow for the testing of further aspects of a real-world service. For example, it may be useful to deny users access to the service an irregular intervals, regardless of the verification performance. This would allow testing of user reactions at the time of access denial and, in follow-up, of the users' long-term attitude to false rejection by the speaker verification procedure.



## **5. System Performance**

### **5.1. Overview**

The required speech technology for both applications is

- text-dependent speaker verification,
- speaker-independent continuous connected digit recognition
- speaker-independent recognition of certain other words and commands
- 

Cut-through is desirable but would not be essential for the service.

In general -automatic telephone services, should be offered in a multi-lingual version. Due to the limited resources CAVE partners have decided to focus current activities on two languages e.g. Dutch and Swiss German-German. Speech synthesis will only have to be a simple word-level concatenative approach.

### **5.2. Error rate**

#### **5.2.1. Speech recognition**

Speech recognition is a requirement in order to implement speaker verification in the real world system. The percentage of string recognition must be at least 95%.

To reach this level of performance, the recogniser will be programmed to perform several checksums on the received cardnumber.

The precise vocabulary of the command words will depend on the exact form and functionality of the application.

#### **5.2.2. Speaker verification**

As regards error rate, clearly, no error rate for speaker verification is too low. Informal discussions and formal interviews so far have pointed to an equal error rate (EER) of 1% as being the maximum which would be acceptable for a real application accessible to the public.

The operational false acceptance and false rejection rates would be set to minimise false rejection to a "tolerable level". Since the best reported EER on telephone speech is around 2% and 5% is generally considered achievable, a figure of 3.5-4% would be acceptable for the banking demonstrator.

#### **5.2.3. Response Time**

Both the speech recognition and speaker verification system must respond within 0.5 seconds with result.

There are two differing response times which should be considered here. The first is the time between enrolment into the system and first use of it. It should be possible to use the system, at the latest, the day after enrolment. This question also ties in with the number of enrolment sessions that will be necessary for use of the service from a variety of different telephones. For the banking application, it is likely that most calls to the service will come either from a "work" telephone or a "home" telephone. This differs significantly from the

---

## Caller Verification in Banking & Telecommunications

**Project** : LE1-1930  
**Deliverable** : Initial Functional Specifications (first phase)  
**Reference** : D 3.1 - public version

---



calling card application, in which calls can come into the system from any telephone. (The CAVE project is in general not considering the problem of calls from the mobile network)

The second response time which must be considered is the time, in the access function, between utterance of the password and the report to the caller of the result of the verification step. This should certainly be no longer than 5 seconds, regardless of the load on the service.

## 6. System Integration

Test versions of the demonstrators will be build and tested in the Netherlands and in Switzerland. Depending on actual results upscaling to large-scale pilots might be discussed.