ODYSSEY04 - The Speaker and
Language Recognition Workshop
Toledo, Spain
May 31 - June 3, 2004

User verification in a BioVXML framework

Enrique Argones Rúa, Elisardo González Agulla,

Carmen García Mateo, Óscar William Márquez Flórez

Signal Processing Group
Signal Theory & Communications Dept
University of Vigo, Spain

eargones@gts.tsc.uvigo.es, eli@gts.tsc.uvigo.es,
Carmen@gts.tsc.uvigo.es, omarquez@tsc.uvigo.es

Abstract

We present BioVXML as an extension of the Voice Extensible Markup Language (VoiceXML). BioVXML is designed for creating modules of biometric verification while maintaining all the features and capabilities of the original VoiceXML.

The user verification can now be performed by processing different biometric traits, such as speech or face, and these features gives us an assurance of the user identity, whereas traditional methods such as written passwords do not. BioVXML provides facilities to develop any monomodal or multimodal biometric recognition system, and it simplifies the creation of remote access control to information systems.

We also show how BioVXML can be used in a practical system like an e-learning platform.

1. Introduction

The design of identity recognition and verification systems has being recently receiving a lot of attention. The advance of the technology in the automatic identity recognition and verification systems has attracted the interest of the industry, and it is possible to make applications that use the increasingly secure biometric accesses.

Classical techniques based on magnetic cards, passwords, electronic or traditional keys have important drawbacks. They can be lost, duplicated, and a cheating user can break the traditional security systems, that actually can't assure the identity of the user. Biometrics can provide a solution to this problem. Physiological features such as fingerprints, iris, voice, face and hand geometry have non-intrusive acquisition techniques and they have the inherent advantages of biometrics because they are own to each user and with an adequate fusion strategy cheats are more difficult.

VoiceXML [1] is a markup language used to define the human-computer dialogue based on voice and other telephone interactions. The use of VoiceXML has increased considerably in recent times. It simplifies the creation of new applications and services based on voice over telephone. By allowing the separation of services from generic media processing, VoiceXML lets developers create new services based on voice without the need of background in voice processing. It allows us to manage audio data in man-machine dialogue systems.

We have extended VoiceXML to adapt it to the identity verification task based on multimodal biometrics. The extension is completely compatible with VoiceXML 1.0 but two new tags have been added and one more has been extended. We have decided to name it as **BioVXML**.

We have put together a demonstration system to show how to use this extension. This demonstration system is the biometric identity verification of the e-learning management system ILIAS [2].

ILIAS consists of tools for learning and teaching on the Internet. One of its tools is an online service to evaluate the acquired knowledge of the students in the different courses. The original purpose of its multimodal identity verification module is to provide this exam tool with the necessary biometric identity verification.

In the following sections we describe the BioVXML specification and an example of BioVXML application where we implement biometric speaker identity verification.

2. BioVXML specification

In the early design phase of the verification module some alternatives were analyzed. Standards such as BioAPI [3], VoiceXML and X+V [4] have been studied, but we have finally opted for making an extension of VoiceXML 1.0 due to its flexibility and an convenient client/server approach.

VoiceXML is originally intended to do applications where the access is usually done by telephone and it just manages audio data. Our application is over a IP network and it must manage multiple kinds of multimedia data. In order to take these aspects into account the VoiceXML must be extended appropriately.

We need a simple way to define human-computer dialogues, but these dialogues must be able to make all the biometric identity verification tasks. BioVXML must be able to handle the different kinds of biometric data, such as voice and face images. Therefore we added the new tags **enroll** and **verify** to the *VoiceXML.dtd*, and we modified the tag **record**. We named the DTD obtained so as *BioVXML.dtd*.

Below we detail the functionality and syntax of the new or modified tags.

- Tag **<record>**: This tag is used for recording audio in VoiceXML. We added the new attribute **src** to it. It indicates the kind of biometric data to be recorded. The DTD element corresponding to **record** will remain:

Table 1: segment of DTD corresponding to the tag record

```
<!ELEMENT record
(%audio;|%event.handler;|filled|grammar\
\|prompt|property)*>

<!ATTLIST record
  %item.attrs;
  type          CDATA          #IMPLIED
  src           CDATA          #IMPLIED
  beep          %boolean;      'false'
  maxtime       %duration;     #IMPLIED
  modal         %boolean;      'true'
  finalsilence     %duration;  #IMPLIED
  dtmfterm      %boolean;      'true' >
```

- Tag **<enroll>**: This new tag is used to make the enrollment task. Its attributes are:
  - **name**: name of the variable associated with the result of the enrollment. Its possible values are *correct* and *incorrect* (*positive* if the enrollment process finishes correctly and *negative* otherwise).
  - **type**: it is used to indicate the kind of enrollment we want to use. This is used to distinguish the different enrollment algorithms from the other ones.

The biometric data are added like parameters. This tag is reflected in the *BioVXML.dtd*:

Table 2: segment of DTD corresponding to the tag enroll

```
<!ELEMENT enroll
(%audio;  |  %event.handler;  |  filled  |
grammar | prompt | param)* >

<!ATTLIST enroll
      %item.attrs;
      type    CDATA    #REQUIRED >
```

- Tag <**verify**>: This new tag is used to make the biometric identity verification. It has the same attributes than the tag enroll (**name** and **type**), with the same meaning. The only difference in this sense is that the possible values of **name** are now *positive*, *negative* or *error* (if an error happens in the verification process). Biometric data are attached as parameters. The *BioVXML.dtd* is modified with:

Table 3: segment of DTD corresponding to the tag verify

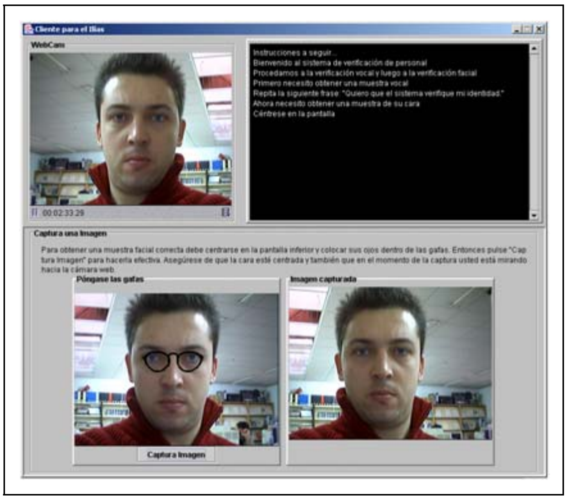
```
<!ELEMENT verify
(%audio; | %event.handler; | filled |
grammar | prompt | param)* >

<!ATTLIST verify
        %item.attrs;
        type           CDATA
#REQUIRED>
```

A parser BioVXML is being developed to interpret the BioVXML documents. These documents must abide by the *BioVXML.dtd* specified. In the present stage of the project, our parser BioVXML allows the record attribute **src** to have two different values, *voice* and *face*, corresponding to audio samples or frontal face images. The parser could implement in a future any biometric value, just adding the associated functionality to the parser. If the attribute src is omitted, we will regard it as an audio source, keeping compatibility with VoiceXML.

3. Example of BioVXML application

The application we developed to demonstrate the functionality and capabilities of BioVXML is a web application with the client/server architecture shown in. It constitutes the biometric identity verification module of the e-learning management system ILIAS. This module offers biometric enrollment and verification methods by using BioVXML dialogues.

The main aim of this application is to test if BioVXML is an adequate framework to make biometric identity verification. This application is an example of implementation and usage of BioVXML.



Instrucciones a seguir...
Bienvenido al sistema de verificación de personal
Procedamos a la verificación vocal y luego a la verificación facial
Primero necesito obtener una muestra vocal
Repita la siguiente frase: "Quiero que el sistema verifique mi identidad."
Ahora necesito obtener una muestra de su cara
Céntrese en la pantalla

Para obtener una muestra facial correcta debe centrarse en la pantalla inferior y colocar sus ojos dentro de las gafas. Entonces pulse "Captura Imagen" para hacerla efectiva. Asegúrese de que la cara esté centrada y también que en el momento de la captura usted está mirando hacia la cámara web.

Figure 1: Frontal face image acquisition interface

The client must handle the different multimedia devices of the computer (i.e. webcam and microphone) in order to get the different biometric data, it must present an adequate user interface and it must send the biometric samples to the server. The solution we finally adopted is a Java applet client in order to get multiplatform compatibility. This solution is easy to integrate into the ILIAS web environment.

Furthermore the server must interpret the BioVXML dialogues and send the adequate commands to the clients. It coordinates the BioVXML dialogues, sending prompts and requesting any record to the client. The server must call the enrollment and verification methods. This allows the dialogue programmers to abstract from biometric algorithmic problems.

Server and client communicate by IP sockets. An ad hoc communication protocol was defined to coordinate this communication. Control and data messages are exchanged by means of this protocol.

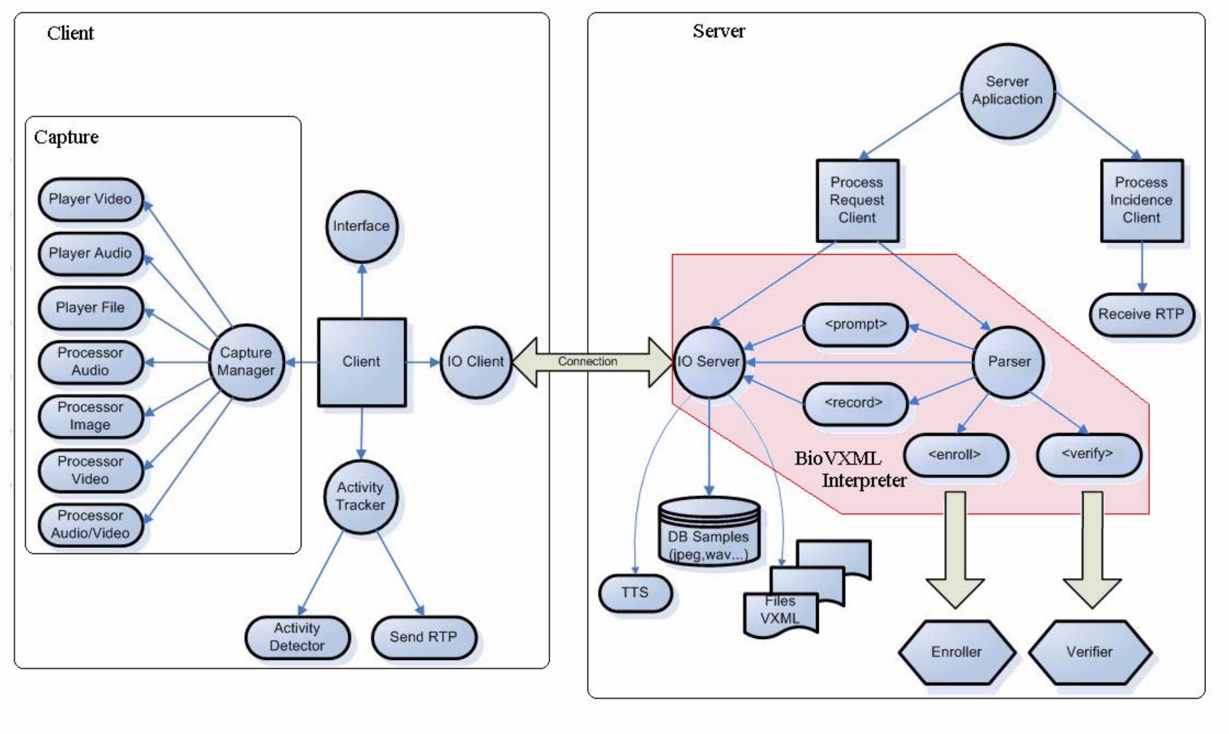


Figure 2: Client/server architecture of the test application

We show an example screen of this application in Figure 1, where we can see the user graphical interface for the frontal face data acquisition.

As a first application we decided to implement a database acquisition module and a verification system based on frontal face images verification and speaker verification. In order to fully deploy a verification application for ILIAS, the stages of the project are the following:

- A biometric database must be designed to tune adequately the different biometric verifiers to use in ILIAS. The name of this database is BioDB, and it has been acquired using the appropriate BioVXML documents in the web application described above.

- Two different biometric verifiers are implemented. The first of them is a speaker verifier, based on GMM. The basis of this speaker verifier algorithm can be found in [5] and [6]. The second one is a multimodal verifier based on the fusion of the speaker verifier and a frontal face verifier based on Bunch Graph Matching (BGM).

4. The Bio-Database (BioDB)

The biometric enrollment task enables us to build an online biometric database. This database will be used to tune the verification algorithms (now and in further works) based on voice and frontal face image and to develop new algorithms. Therefore we must collect the Bio-Database taking into account the needs of the different biometric verification algorithms.

We have decided that in an application like the one Ilias defines, speech and face are the most adequate biometric traits to be processed. So, our goal for the first biometric verification system is to develop a multimodal system that jointly uses speech and face verification, hence the BioDB must have frontal face images and speech samples from every user in order to train and test expert and fusion systems.

On the other hand, the BioDB should be an open set database, like the BANCA database, and therefore it must be designed in a similar way.

Each session consists of ten sentences and two frontal face images taken at the beginning and at the end of the speech sentences. These sentences are:

- A set of six phonetically balanced sentences (in Spanish).

- The user’s full name.

- Other user’s full name (attack to that user).

- The user’s pin.

- The other user’s pin (attack to that user).

The photos are taken at the beginning and at the end of the session. The users are divided in groups. The users of a group will do an attack to every other member of that group, so that every user is attacked the same number of times, and every user has the same number of true and false identity claims.

The database is still in a draft stage. The number of users is still too low, and the number of male and female members is unbalanced. There are 12 users, divided in two groups of 6 users each. There is just one female in each group. Hence every user in the database has recorded five sessions.

The environment in the acquisition process is a realistic one. The users have recorded their sessions in their usual offices, where there is an important background noise and the illumination is not controlled.

The acquisition hardware that we have used consists of several low-range web-cams such as the NGS Showcam Plus, Creative WebCam NX Pro, Philips ToUcam PRO II and the Logitech QuickCam Zoom. The microphones used are the internal web-cam microphone or a poor quality microphone if the camera does not incorporates one.

The quality of the sound and the face images is the one we can expect in an internet application.

The experimental protocol used in our preliminary experiments is the following:

1. Choose one group as the *development group* and the other as the *test group*.
2. Train the monomodal expert using the sessions 1 and 2 (in the *development* and in the *test* groups).
3. Use the sessions 3, 4 and 5 of the *development group* to tune the expert thresholds or the fusion algorithms.
4. Use the sessions 4, 5 and 6 of the *test group* to evaluate the performance of the system.
5. Repeat from 1. inverting the group selection.
6. Average the results obtained in 4 in both iteractions.

Neertheless this database will grow to enable us to do an adequate gender group separation, to build bigger groups to make the generalization easier. The actual conditions are too hard for the state-of-art algorithms. To facilitate this, users and any voluntary contributor can access the biometric verification module to enroll in the BioDB or to verify its identity in [9].

5. Description of the implemented biometric verifiers

We have developed two verification systems: a text-independent speaker verifier and a multimodal biometric verifier.

The text-independent speaker verifier uses Gaussian Mixture Models on the MFCC vectors. A universal speaker model has been previously trained with Spanish speakers data. We made a Maximum Likelihood Linear Regression (MLLR) to adapt the universal model to every client, and so we get the client models. With this approach the problem of sparseness of client audio data is avoided. The number of gaussian mixtures we have chosen for the universal speaker model is 128 whereas the user models are obtained adapting a 64 universal mixture model.

The verification process is based on comparing the normalized score of the speech received using the claimed user model with a global threshold. The universal 128 mixture model and the duration of the speech are taken into account in order to do the score normalization.

On the other hand, the multimodal biometric verifier uses the expert described above as speaker verifier and it is fused with a frontal face verifier based on Bunch Graph Matching.

The operation of these verification systems can be tested in [9] and in a demonstration during this workshop.

6. Conclusions and further work

We have described BioVXML specifications. BioVXML allows us to integrate quickly and easily new biometric algorithms in the application. The definition of the dialogues is easy and intuitive, and the compatibility of BioVXML with VoiceXML 1.0 makes possible the integration of BioVXML dialogues in complex VoiceXML systems.

BioVXML can be easily used to do a large variety of biometric tasks, such as doing a biometric application enrollment, acquiring a database or verifying the identity of a user, and it can be easily used in web applications, such as the e-learning platform ILIAS.

We have collected a biometric database, BioDB. This database is intended to be a tool to adjust the different biometric algorithms in the web environment. It is in a draft state, but its open-set structure allows us to create new groups of users and to add people to the existing ones, making the growth of the number of users easier.

On the other hand, the acquisition environment is hostile enough to any state-of-art biometric identity verification system based on face images or voice. Low-range hardware has been used to get the biometric samples, and noise and illumination are not controlled, neither in training sessions nor in test sessions. These factors make the BioDB an adequate database to tune web oriented biometric algorithms.

7. Acknowledgements

This project has been partially supported by Spanish MCyT and Xunta de Galicia under the projects TIC2002-02208 and PGIDT02TIC32201PR respectively.
We would like to thank Daniel González Jiménez for his collaboration developing the BGM image verifier.

8. References

[1] VoiceXML 1.0 <http://www.w3.org/TR/voicexml/> 2003, 22 September

[2] Url del Ilias. ILIAS open source documentation <http://www.homer.ilias.uni-koeln.de/iliasdoc/doc/html/1.html>

[3] BioAPI Consortium , BioAPI Specification Version 1.1 <http://www.bioapi.org> 2003, 15 September

[4] X+V 1.1 --- XHTML + Voice Profile <http://www.voicexml.org/specs/multimodal/x+v/11/> 2003, 2 October

[5] Enrique Argones-Rúa, Daniel González-Jiménez, José L. Alba Castro, Carmen García-Mateo, *Hierarchical Multi-Modal Fusion for Identity Verification*, submitted to the International Conference on Biometrics Authentication, Hong Kong, 2004.

[6] Leandro Rodríguez-Liñares, Carmen García-Mateo, José Luis Alba-Castro, *On combining classifiers for speaker authentication*, Pattern Recognition Journal (Elsevier Science). vol. 36 , pp.347-359. February 2003.

[7] Bailly-Baillire E., Bengio S. et al., *The BANCA Database and Evaluation Protocol*, AVBPA 2003.

[8] S. Ben-Yacoub, Y. Abdeljaoued, E. Mayoraz, *Fusion of Face and Speech Data for Person Identity Verification*, IDIAP Research Report, January 1999.

[9] Prototype of the ILIAS biometric user identity verification module, <http://leirado.gts.tsc.uvigo.es/~ILIAS/>, January 2004.