



# **Speaker Identification and Verification (SIV) Introduction and Best Practices Document**

Internal Working Draft – February 13, 2006

**VoiceXML Forum  
Speaker Biometrics Committee**

Authors:  
Valene Skerpac, iBiometrics, Inc. (iBICS)

## About the VoiceXML Forum

Voice Extensible Markup Language (VoiceXML) is a markup language for creating voice user interfaces that use automatic speech recognition (ASR) and text-to-speech synthesis (TTS). Since its founding in March 1999, the VoiceXML Forum has continued to develop, promote and to accelerate the adoption of VoiceXML-based technologies via more than 150 member organizations worldwide.

Tens of thousands of commercial VoiceXML-based speech applications have been deployed across a diverse set of industries, including financial services, government, insurance, retail, telecommunications, transportation, travel and hospitality. Millions of calls are answered by VoiceXML applications every day.

The Forum's primary focus areas include:

- Promoting the adoption of VoiceXML-based technologies
- Cultivating a global VoiceXML ecosystem
- Actively supporting standards bodies and industry consortia, such as the W3C and IETF, as they work on VoiceXML and related standards, such as CCXML, X+V, MRCP, and speech biometrics.

For more information on the VoiceXML Forum visit the website at <http://www.voicexml.org>.

## Disclaimers

This document is subject to change without notice and may be updated, replaced or made obsolete by other documents at any time.

The VoiceXML Forum disclaims any and all warranties, whether express or implied, including (without limitation) any implied warranties of merchantability or fitness for a particular purpose.

The descriptions contained herein do not imply the granting of licenses to make, use, sell, license or otherwise transfer any technology required to implement systems or components conforming to this specification. The VoiceXML Forum, and its member companies, makes no representation on technology described in this specification regarding existing or future patent rights, copyrights, trademarks, trade secrets or other proprietary rights.

By submitting information to the VoiceXML Forum, and its member companies, including but not limited to technical information, you agree that the submitted information does not contain any confidential or proprietary information, and that the VoiceXML Forum may use the submitted information without any restrictions or limitations.

## Revision History

Date	Description
February 13, 2006	Internal Working draft – Introduction section for review, Outline of Best Practices section for review

## INTRODUCTION AND BEST PRACTICES

### TABLE OF CONTENTS

<b>1.</b>	<b>GOALS OF DOCUMENT</b>	<b>4</b>
<b>2.</b>	<b>OVERVIEW OF TECHNOLOGY</b>	<b>4</b>
2.1.	Automatic Biometric Processing for Information Security	4
2.2.	Speaker Identification and Verification Technology	6
2.3.	Positioning of Speaker Identification and Verification Technology to VoiceXML SIV	9
<b>3.</b>	<b>APPLICATIONS (OUTLINE FORM)</b>	<b>10</b>
3.1	Feasibility Assessment	
3.2	Life Cycle Approach	
3.3	User Interface	
3.4	Other Identifiers Used	
<b>4.</b>	<b>VOICE ENGINE (S) MANAGEMENT (OUTLINE FORM)</b>	<b>10</b>
<b>5.</b>	<b>SECURITY (OUTLINE FORM)</b>	<b>10</b>
5.1	Core Management and Security Requirements required for an SIV applications	
5.2	Architectures Covered	
5.3.	Assessments	
5.4.	Controls in the environment	
5.5.	Key Management Life Cycle Controls	
5.6.	Biometric Information Life Cycle Controls	
5.7.	Security Examples	
5.8.	Privacy	
<b>6.</b>	<b>REFERENCES</b>	<b>14</b>
<b>7.</b>	<b>TERMINOLOGY</b>	<b>14</b>

## 1. Goals of the document

This document is intended to complement SIV Specifications and be used in conjunction with Voice XML Application, Architecture and Data Interchange documents under development. Its approach to biometric processing and security discussions is based on the X9.84-2003 ANSI standard, Biometric Information Management and Security for the Financial Services Industry (Ref. 6.1). Unlike X9.84, biometric objects and encoding specifications are beyond the scope of this document (*note: would like to say where these are addressed in SIV documentation*).

The SIV Introduction and Best Practices document provides guidance to entities that plan to implement Voice XML SIV applications that perform authentication with a certain degree of accuracy during conversations. This capability represents a shift in today's typical biometric authentication model towards a more granular application usage. To further the general acceptance and success of SIV applications, we look to utilize mature technologies and apply current best practices in areas such as biometrics, voice systems, security and privacy.

## 2. Overview of Technology

User authentication, which confirms the identity of an individual, is a counter measure against computer security exposures. Examples of user authentication include identification of the sender of a message or verification of the identity of the initiator of an electronic transaction, or a document author. Biometrics statistically measures certain human anatomical and physiological traits that are unique to an individual. Information Security recognizes biometrics as an authentication method that identifies or verifies user-characteristics representing 'who the user is'. Other user authentication methods recognized are 'something in the user's possession' such as a smart card or 'something the user knows' such as a password.

Characteristics used as biometric representations of the user today include voice as described in section 1.2, fingerprints, hand geometry, iris patterns, retinal patterns, facial image, and signature verification as well as others. Automated authentication of each biometric has its own set of properties to be considered for each application. Properties include public perception and policy, level of fraud resistance, comprehensiveness, uniqueness, accuracy (match rates and error rates), degree of permanence, storage space, performance, capabilities for system validation, environmental and interface factors.

### 2.1 Automatic Biometric Processing for Information Security

Biometric processing consists of the automatic capture and comparison of a biometric characteristic. The digital representation of the characteristic produced is electronically stored for subsequent validation of the user's identity. The following four basic steps are involved in biometric authentication:

- Input of the biometric
- Quality analysis and potential re-capture of the biometric input
- Creation of digital representation of the captured biometric
- Match digital representation with previously enrolled representation(s) to determine if a match exists.

The three biometric processes associated with biometric authentication are:

**Enrollment:** The process of gathering biometric samples from a user and generating and storing biometric reference models for the individual. Enrollment can involve the collection of other information about the user establishing organization, account and user privileges. Enrollment can be preceded by various searches, including a 'one-to-many' biometric comparison, to insure that the user is not already enrolled in the database.

**Verification:** Verification confirms that the user is who he claims to be by performing a 'one-to-one' comparison of the enrolled biometric reference model to the newly captured sample.

**Identification:** The identification process identifies a user against a database of enrolled biometric references by performing a 'one-to-many' comparison to the newly captured biometric sample.

Though biometric characteristics, applications and specific methods of biometric authentication vary widely, **a generic biometric system model** is recognized for the purposes of standards. This document focuses primarily on SIV systems but recognizes that it fits within the biometric architecture described below. The architecture's major components are referenced throughout this document and in particular, section 5.3.2, Threat Assessments and Protection Methods. The advantages to referencing the generic biometric architecture are to utilize previous efforts and knowledge developed from existing biometrics standards documents and to facilitate communication between biometric standards and knowledgeable professionals fostering a better understanding of SIV.

Major Components of a generalized biometric architecture are:

- Data Collection
- Signal Processing (Feature Extraction)
- Matching
- Decision
- Storage
- Transmission

**Data Collection:** The data collection component consists of an input device that captures biometric information from the user and converts it to a form for processing. It links the physical environment to the logical domain. Its output is considered the raw biometric data.

**Signal Processing:** The signal processor receives the raw biometric data from the data collection subsystem and converts it to a form required by the matching component. The signal can be filtered to remove noise or other extraneous data to the matching process and may be normalized in some way. After pre-processing, feature extraction processing creates a digital representation of the characteristics from the raw biometric data, which is to be used by the matching process.

**Matching:** The matching component receives the biometric data from the signal-processing component and compares it with stored biometric models. The following sub-components comprise the matching component:

- A sequencer (controls sequencing of match, adaptation and transfer of scores to decision subsystem);
- A match scoring module (measures similarity of claimant sample with model); and
- An adaptation module (optional)

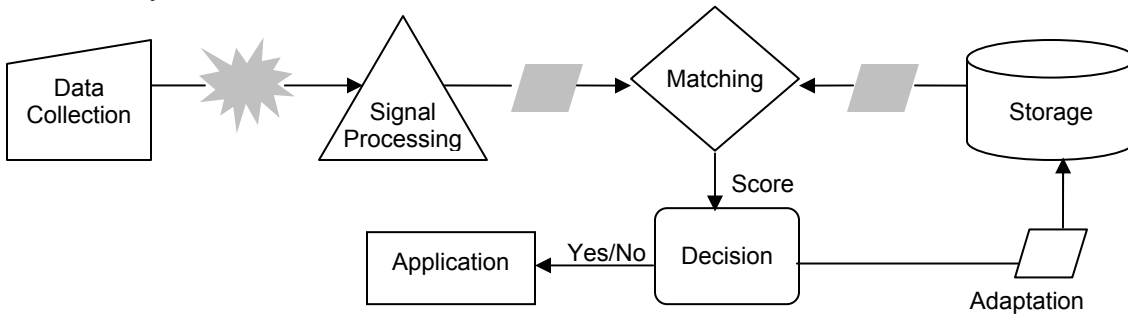
Matching can be a straightforward sequence of events or involve interaction between subcomponents and even feedback from the decision processing depending on the biometric application.

**Decision:** The decision component receives a score from the matching subsystem and assesses the results of the score using a confidence value based on business risk and risk policy. A binary yes or no decision is returned regarding the affirmative identification or verification of the user based on the score result. Often times, a single threshold is used whereby the score must not exceed a prescribed threshold. More involved approaches might be used depending on the application such as a set number of matches on multiple biometric samples or on multiple biometric characteristics.

**Storage:** The storage component maintains enrolled users' biometric models, which includes addition, deletion and retrieval of models as required by the matching component. Models can be stored in a traditional database on a computer system, protected storage of biometrics device, or on a portable tokens, such as a smart card. In addition to the users' biometric model, other information and unrelated data could be stored on the database.

**Transmission:** The transmission component sends information between the collection, signal process, decision, storage and matching components. Connectivity can be from one point to another or networked where one system connects to multiple components. System components can be local or remote to each other using the same or separate security techniques.

The following, **Figure 1**, illustrates the verification process using the major components of the biometrics system model:



**Figure 1** *Generic Biometric Model*

## 2.2. Speaker Identification and Verification Technology

The following section references the preceding biometrics sections and focuses on automatic speaker recognition as a biometric for information security. It provides background information on SIV technology based on 'An Overview of Automatic Speaker Recognition Technology' by Douglas A. Reynolds of MIT Lincoln Laboratory (Ref 6.2).

Automatic speaker recognition systems extract, characterize and recognize information from the speech signal, which conveys the speaker's identity. Identity is derived from the shape of the speech spectrum, which encodes information about the speaker's vocal tract shape via resonances and glottal source via pitch harmonics. Consistent with biometric terminology described in section 1.1, *speaker identification* determines who is speaking from a known set of voices whereby no claim of identity is made and a 'one-to-many' comparison is performed. *Speaker verification* determines if the user is whom he/she claims to be resulting in a yes/no decision.

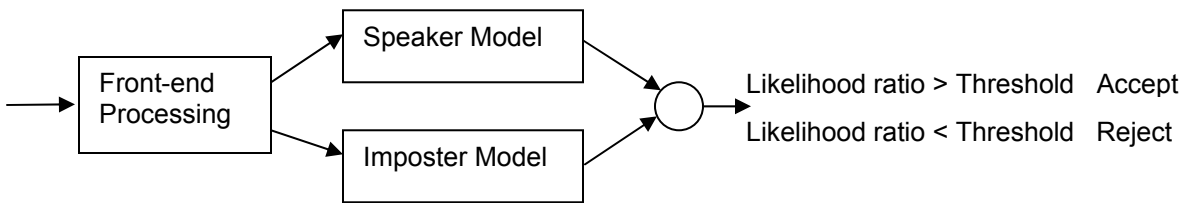
Applications specify the level of cooperation and control by the user, which determines the use of either *text-dependent* or *text-independent speech*. Text-dependant applications have prior knowledge of the text to be spoken and the user cooperatively speaks this text. Text-independent applications have no prior knowledge by the system of the text to be spoken. Processing of text-independent speech is more difficult but also more flexible.

Properties of speaker recognition as a biometric considered compelling include a positive public perception and available technology through a multitude of voice capable devices and networks. (Note: Discuss further other characteristics growing more positive and challenges: degree of fraud resistance, universality, uniqueness, accuracy (match rates and error rates), stability, storage space, performance, capabilities for system validation, environmental/ interface factors.)

**SIV Architecture**

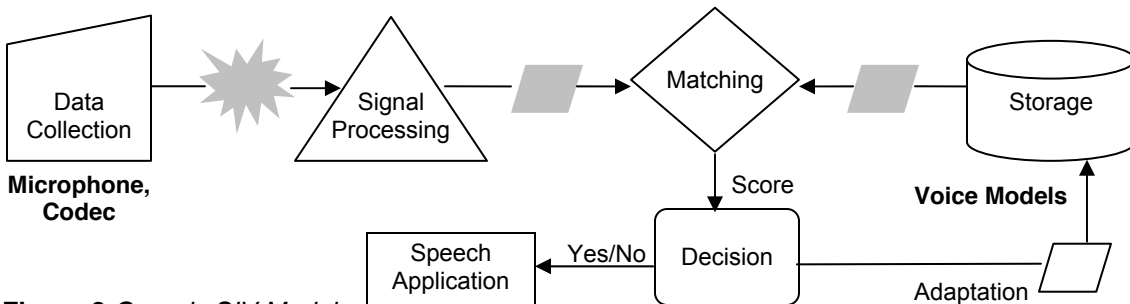
Modern speaker verification systems, as described in the referenced paper and shown in **Figure 2** below, perform a *Likelihood Ratio test* that distinguishes between two assumptions: the speech comes from the claimed speaker or from an imposter. Features extracted from the users speech in the front-end processing are compared to both the claimed speaker model and the potential imposter speaker’s model(s). The Likelihood Ratio is derived by calculating the difference in the match score results and then used to compare to the Threshold.

**Figure 2 Traditional Speaker Verification Model**



**Generic SIV Model**

The following section uses the general biometric system model and traditional speaker verification model to create a generic SIV model, as shown in **Figure 3**, to be used for reference throughout this document. The following figure illustrates a verification (only) process using the major components of the generic SIV system model:



**Figure 3 Generic SIV Model**

Major Components of a generalized **SIV architecture** are:

**Data Collection:** Speech collection is the first piece of Front-End Processing in the traditional speaker verification model. It is performed through a Microphone input device that converts sound waves into analogous electrical waves. The microphone’s basic component consists of a diaphragm that responds to the pressure or particle velocity of sound waves. A Codec samples and encodes the input signal, typically creating a standard or proprietary form of raw speech data (i.e. speech signal). Standard forms, for example include the ITU-T G.711 telephony standard which uses 8 bit pulse code modulation (PCM) samples for signals of voice frequencies, sampled at the rate of 8000 samples/second. A list of audio codecs can be found at [http://en.wikipedia.org/wiki/List\\_of\\_codecs](http://en.wikipedia.org/wiki/List_of_codecs).

**Signal Processing:** Signal processing is the second piece of Front-End Processing in the traditional speaker verification model and it comprises of three parts. One is the detection of speech from the raw speech data and the filtration of non-speech. The second part is the extraction of features that convey speaker information from the filtered speech. Feature extraction typically applies short-term analysis with 20 ms windows placed every 10 ms to

compute a sequence of measurements using a number of techniques. This data is then converted to specific features via various methods.

The third part of signal processing is channel compensation, which diminishes the effects of the input device by applying adjustment to features. Other methods to remove channel effects are possible in the matching component as well.

**Matching:** During enrollment, speech is collected and features are used to generate a voice model<sup>1</sup> that is representative of the speaker. There are a number of modeling techniques used to create an appropriate voice model<sup>2</sup>. Imposter models can be crucial to optimal performance acting primarily as a normalization to help minimize non-speaker related variability in the likelihood ratio score. Selection of modeling and related techniques is dependent on the type of speech, anticipated performance, ease of training and updating and storage and computation requirements.

Speech pattern matching computes a score, which measures how similar the input features are to the voice model. Speaker adaptation, which updates the voice model to better represent the user, can occur during the matching process.

**Decision:** As a result of SIV score matching, a decision to accept, time-out, request for more speech or reject is made. As shown in the proceeding SIV architecture section, the score matching process leads to a Likelihood Ratio, which is compared to a Threshold to decide to accept or reject. Various methods to determine an appropriate Threshold include a minimum error performance between real and imposter speaker, a fixed False Match Rate (also known as False Acceptance (FA)) or False Non-Match Rate (also known as False Rejection (FR)) criterion, and a desired FA/FR ratio.

**Storage:** The storage component maintains enrolled users' voice models, which includes their addition, deletion and retrieval as required by the matching component. Voice models are traditionally stored in a protected central database but can be stored on protected portable tokens, such as a smart card.

**Transmission:** The transmission component sends information between the collection, signal process, decision, storage and matching components. Speech data collection and signal processing can be performed locally or remotely through networks, which include the legacy circuit switched networks, cellular networks and Voice over Internet Protocol (VoIP) networks. Depending on the many potential configurations, speech can be carried via analog or digital signals.

Recent availability of advanced converged voice and data platforms and networks expands the traditional telephony speech model, which assumes data collection and signal processing over a legacy network and voice matching, decision, storage and application control on a protected central processor. Reflective of this diversity, transmission of biometric speech data varies based upon configuration.

---

<sup>1</sup> Other terms Speaker Model or Voice Print are sometimes used to refer to the voice model.

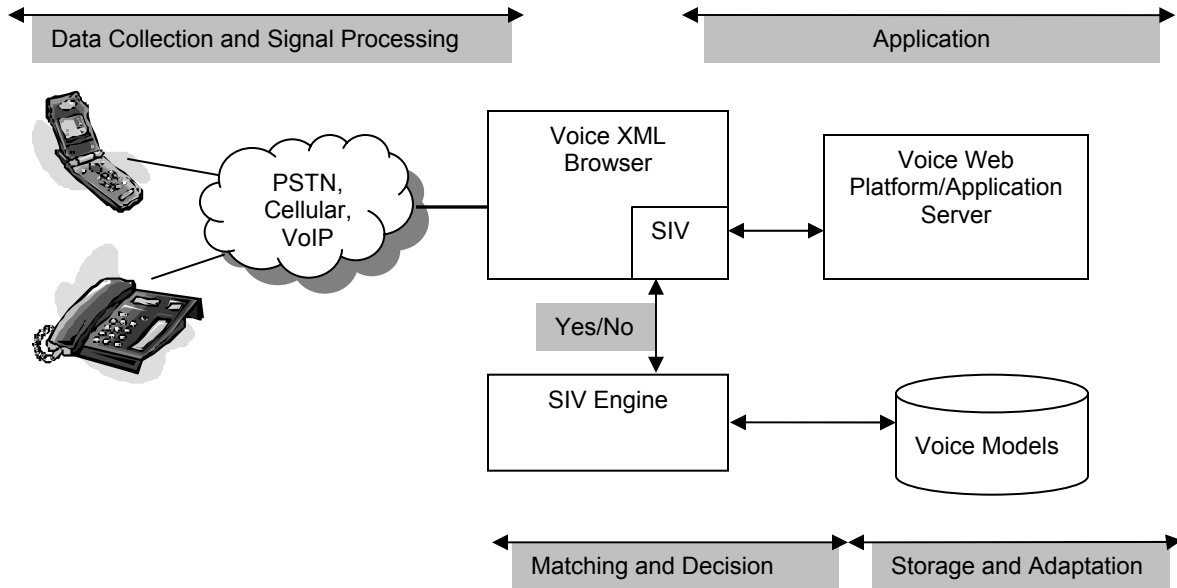
<sup>2</sup> Methods include Template matching, Nearest neighbor, Hidden Markov Models (HMMs) and Neural Networks, see reference for more information on modeling techniques.



**2.3. Positioning of Speaker Identification and Verification Technology to VoiceXML SIV**

The following section references the preceding biometrics and SIV sections and relates it to the developing SIV specifications for the Internet, specifically those from the VoiceXML Forum and WC3 standards organizations.

VoiceXML standardizes SIV for web oriented speech applications. It incorporates all three biometric processes (i.e. enrollment, verification, and identification) and supports the components of the generic SIV model (i.e. data collection, signal processing, matching, decision, storage and transmission). Figure 4 loosely maps the VoiceXML SIV architecture to the major components of the generic SIV model.



**Figure 4** VoiceXML SIV Architecture mapped to SIV Model Components

VoiceXML SIV architecture anticipates multiple speech-enabled device types with external or imbedded microphones and codecs. It assumes Data Collection and Signal Processing over a variety of public and private networks with any number of devices. Matching and Decision Processing is performed through VoiceXML SIV, which operates in concert with the application and utilizes voice engine resources. Storage and Adaptation Processing is performed by the voice engine resources and controlled by the SIV application.

In a VoiceXML environment, Transmission is normally managed distinctively in two parts, one part being the front-end, which comprises of the data collection and signal processing components and two being the back-end, which consists of the other components. Evolving converged digital technologies can enable end-to-end management of voice, which is an important aspect of new speech technologies. This Best Practices document will address various environments with an emphasis on those most widely utilized.

## **SIV Best Practices - Outline**

This Best Practices section intends to discuss application, management, security and privacy areas of VoiceXML SIV providing insight into these important aspects of deployment.

### **3.0. Applications**

This section compliments the SIV Applications Document, which contains specific real world and future applications.

#### 3.1 Feasibility Assessment

Where can SIV be used and what are the current limits of the technology.

#### 3.2 Life Cycle Approach

Development phases and timing

'Function Creep' prevention

Monitoring and compliant logging

#### 3.3 User Interface

#### 3.3 Other Identifiers used

Non-biometric factors to consider upon authentication, for example caller-ID

Authentication outcome calculation, weighting for each factor

### **4.0. SIV Voice Engine(s) Management**

#### 4.1. Threshold management

#### 4.2. Accuracy tuning

#### 4.3. Performance evaluation

### **5.0 Security**

#### **5.1 Core Management and Security Requirements required for an SIV application.**

##### 5.1.1. Enrollment

##### 5.1.2. Verification

##### 5.1.3. Identification

##### 5.1.4. Adaptation

##### 5.1.5. Transmission and Storage

##### 5.1.6. Termination and Archive

##### 5.1.7. Audit Journal

#### **5.2 Architectures Covered**

##### 5.1.1. Central Reference Model

##### 5.2.2. Distributed Reference Model (Does SIV architecture support imbedded?)

**5.3. Assessments**

5.3.1. Risk Assessment, Legal and Regulatory requirements

5.3.2. Threat Assessments and Protection Methods

This section describes security consideration and potential attacks or vulnerabilities in voice biometric systems (based on X9.84 biometrics standards approach but made specific for speech systems).

a. False identify of registered user

Exposure Point	Methods of Protection	Component Vulnerable
1. Enrollment Process	1. Clearly articulated and controlled enrollment	<ul style="list-style-type: none"> <li>• Capture</li> <li>• Process</li> <li>• Transmission</li> <li>• Storage</li> </ul>

b. Fraudulent Voice Input

Exposure Point	Methods of Protection	Component Vulnerable
1. Methods for collecting voice which can be used to launch a fraudulent voice attack include: <ul style="list-style-type: none"> <li>a. Voice from pre-recorded file</li> <li>b. Raw voice data or reference model obtained from system (wiretaps, database access, etc)</li> <li>c. Fake station, web site or other false authority that users think are real that obtain raw voice data or reference model.</li> </ul>	1. Engine detection of recordings 2. Variation of prescribed input to ensure 'liveness'. 3. Authentication of authority through trusted third parties 4. Monitoring of remote station, camera, live attendant, etc,	<ul style="list-style-type: none"> <li>• Capture</li> </ul>

c. Protection of data

Overview

Replay attack – voice insertion into system

Search for chosen reference model similar enough to verify as other person

Search for two reference models similar enough that one person can collude in fraud

d. Verification or identification result is changed

e. False match versus false non-match

Overview

Improper Threshold Settings

Improper Tuning

Illicit system performance

f. Scores and Thresholds

Hillclimbing attack

Update and adaptation

g. Single versus multi-factor authentication

h. Testing

i. Open versus closed voice system

j. Compromise/loss of biometric data

k. Compression

l. System circumvention

**5.4. Controls in the environment**

*Sample of How to Organize Controls section of SIV Best Practices Document*

This sample is intended as an example of how to organize the control section for an SIV environment. The following is provided as an example from X9.84, which is an ANSI standard currently obtained for a fee and should not be widely distributed. I think it would be fairly straightforward to draw upon the X9.84 Biometric standard control criteria list and apply it to SIV.

Security Policy

	<b>Control Criteria: Information Security Policy</b>	<b>References</b>
1.	A voice biometric information management and security policy document for the organization. Document is approved by management, published and communicated, as appropriate, to all employees, customers and users of the SIV system.	ISO 17799 3.1.1 Etc.
2.	At a minimum, the policy contains the following: A) A brief explanation of the Organizations security policies, principles, standards and compliance requirements of particular importance to the organization including: a. Compliance with legislative and contractual requirements, b. Security education requirements, c. Prevention and detection of viruses and other malicious software d. Cryptography requirements, e. Business continuity management, and f. The consequences of security policy violations B) Etc	Etc
3	Review Process	
	<b>Control Criteria: Policy Management</b>	<b>References</b>
4	The organization publishes the applicable public section of policy to all appropriate users.	etc
	etc	
8		

Organization

	<b>Control Criteria: Information Security Infrastructure</b>	<b>References</b>
9.	A management group or security committee exists to coordinate the implementation of information security measures.	
	<b>Control Criteria: Security of Third Party Access</b>	<b>References</b>
13.	Procedures exist and are followed to control access to organizational information processing facilities by third parties.	etc
	etc	

Policy, Organization, Asset management, personnel security, physical security, operations management, system access management, systems development and maintenance, business continuity, monitoring and compliance, journaling

**5.5. Key Management Life Cycle Controls****5.6. Biometric Information Life Cycle Controls**

## 5.6.2. Enrollment controls

## 5.6.3. Reference Model life cycle

Are there legal constraints for how long or under what circumstances biometric information can be retained? Typically not covered by law, but usually by contracts between users and service providers.

## 5.6.4. Verification and ID process controls

## 5.6.5. Device controls

**5.7. Security Examples****5.8. Privacy**

## 5.8.1. Assessments

5.8.2. *Controls*

## 5.8.3. Methodologies

Security and biometric processing topics are taken from the X9.84-2003 Biometric Security standard that includes architecture and technical specification components within one document. The X9.84 standard accommodates all biometrics and various security levels. The SIV Best Practices document, on the other hand, accommodates various SIV architectural designs, applications and security levels. (I assume the SIV architecture document will be referenced throughout the Best Practices document.)

Though the categories are general for organizational purposes, it is the intension to focus on voice/speech related issues.

## 6. References

1. "Biometric Information Management and Security for the Financial Services Industry", X9.84-2003 ANSI standard, [www.x9.org](http://www.x9.org).
2. "An Overview of Automatic Speaker Recognition Technology", Douglas A. Reynolds, MIT Lincoln Laboratory, MA, USA, [www.ll.mit.edu/ST/pubs/020513\\_Reynolds.pdf](http://www.ll.mit.edu/ST/pubs/020513_Reynolds.pdf), 2002.

## 7. Terminology

This section includes an enumeration of terms and abbreviations related to biometrics technologies and SIV technology as developed in the SIV Requirements document. The purpose of this section is to establish a common list of terms used in the VXML BioSig discussions and within this set of documents.

### **adaptation**

The process of automatically updating or refreshing a reference model.

### **ASR**

Acronym for automated speech recognition

### **attempt**

The submission of a voice sample on the part of an individual for the purposes of enrollment, verification, or identification in an SIV system. A user may be permitted several attempts to enroll, to be verified, or to be identified.

### **authentication**

1. The process of confirming one or more identity claims. Synonymous with verification and identification when a claim of identity has been made.
2. The process of confirming one identity claim. Synonymous with verification.
3. Synonym for authenticity.

### **authenticity**

One of the basic security requirements. Protection of SIV data from generation by an unauthorized source and modification

### **biometric fusion**

When two or more biometrics are used in a single transaction and the results are combined to produce an overall score.

### **buffering**

The result of a pre-processing stage. The nature and content of the resulting buffer vary with SIV engines. Buffering is defined as preprocessing for future SIV processing that can be performed without knowing the identity claims or the type of operation to be performed (i.e. enrollment or authentication).

### **best match speaker**

A part of an identification result that includes the identity claim out of the provided list of claims that the SIV system detected as the speaker that spoke the input speech sample.

### **closed set identification**

Identification performed by a system that does not employ imposter models. The result returned from Closed Set Identification always includes a top-match speaker. Closed set

identification can therefore assume that the top-match speaker exists in the list of claimed identities.

**capture**

The acquisition of a spoken sample.

**challenge-response**

synonym for text prompted

**claim of identity**

The name or index of a claimed reference model or enrollee used for verification.

**claimant**

A person submitting a spoken sample for verification claiming a legitimate or false identity.

**confidentiality**

One of the basic security requirements. Protection of data, including SIV data, from unauthorized access and inadvertent disclosure

**decision policy**

The logic through which an SIV system provides match / no match decisions, inclusive of the following elements:

- The SIV system's matching thresholds
- The number of match attempts permitted per transaction
- The number of reference models enrolled per claimant
- The number of distinct speech samples enrolled per claimant
- Other security factors (e.g., other biometrics, PINs, tokens)
- The use of internal controls in the matching process to detect like or non-like samples
- The use of serial, parallel, weighted, or fusion decision models that utilize more than one reference model in the match process for a given user

**encryption**

A process of transforming plaintext (readable) into ciphertext (unreadable) for the purpose of security or privacy.

**enrollment**

The process of collecting voice samples from a person and the subsequent generation and storage of voice reference models associated with that person. See also initial enrollment and re-enrollment.

**failure to acquire**

Failure of an SIV system to capture a biometric sample, or to extract SIV data from input speech, sufficient to generate a reference model or perform authentication.

**failure to enroll**

Failure of an SIV system to capture one or more voice samples, or to extract SIV data from one or more voice samples, sufficient to generate a reference model.

**false match rate**

In a One-to-One system, the probability that a system will falsely verify an imposter as a legitimate enrollee. In a One-to-Many system, the probability that a system will incorrectly identify an individual. Historically also known as a Type II Error from hypothesis testing. Same as False Acceptance Rate.

**false non-match rate**

In a One-to-One system, the probability that a system will fail to verify the identity of a legitimate enrollee. In a One-to-Many system, the probability that a system will fail to identify a legitimate enrollee. Historically also known as a Type I Error from hypothesis testing. Same as False Rejection rate.

**identification**

Authentication with multiple identity claims. An identification result includes both the verification results for all of the individual identity claims, and the identifier of a single reference model that matches the input utterance best. The single best match result may be blank, indicating that no user has been identified. Same as – one-to-many, multi-verification.

**impostor**

A person who submits a voice sample in either an intentional or inadvertent attempt to be verified or identified as another person who is an enrollee.

**impostor models**

One or more models used by an SIV as an internal description of the counter hypothesis that the input utterance was spoken by one of the claimant speakers. Impostor models are used to model the out-of-set hypothesis and enable a system to perform Open Set Identification.

**initial enrollment**

The process of enrolling an individual's voice data for the first time, such that the individual must provide a non-biometric means of authentication such as a password or ID in order to establish or confirm an identity. Enrollment results in a reference model. See also enrollment and re-enrollment.

**integrity**

One of the basic security requirements. Protection of SIV data from undetectable modification and substitution.

**match**

The process of comparing a match model against a previously stored reference model, and scoring the degree of similarity or correlation between the two. Authentication comprises of performing match operations followed by a decision policy.

**multi-biometric authentication**

Authentication using two or more different biometric types, for example:  
finger biometric with iris biometric, voice biometric with face biometric.

**multi-factor authentication**

Multi-factor Authentication is the combination of two or more authentication techniques that together form a stronger or more reliable level of authentication. This usually involves combining two or more of the following types:

- Knowledge factor, "something an individual knows"
- Possession factor, "something an individual has"



- Biometric factor, “something an individual is”
- Location factor, “where you are” **Added**

**non-repudiation**

Protection of SIV data from renunciation and denial of issuance

**one-to-many**

See Identification.

**one-to-one**

See Verification.

**open set identification**

Identification that may return a blank best-match speaker in its result, implying that the underlying SIV system employs imposter models. This works assumes that identification is always open set.

**privacy**

The right of an individual, group, or institution, to control, edit, manage, and delete information about themselves and decide when, how, and to what extent that information is communicated to others.

**raw voice data**

The captured, unprocessed voice data in digital form suitable for subsequent SIV processing.

**re-enrollment**

The process of enrolling an individual’s voice data where the same or other voice data have been enrolled at least once. See also enrollment and initial enrollment. Same as Adaptation.

**reference model**

Data that represents the voice measurement of an enrollee. It is based on data extracted from one or more voice samples provided by that individual and is typically stored and used by an SIV system for comparison against subsequent submitted voice samples. Also see voice model and voiceprint.

**replay attack**

The use of the tape recorder or other recording device to record verification or enrollment utterances that are then used to spoof and SIV system.

**score**

A numerical representation of the degree of similarity between an input speech sample and a reference model. The specific method by which a score is generated, as well as the probability of its correctly indicating a true or false match, is generally propriety to each vendor.

**scoring**

Synonym for authentication (since authentication results include a score).

**single factor authentication**

Authentication using only one identity factor. Also see multi-factor authentication

**SIV**

Acronym for speaker identification and verification

**SIV data**

Extracted information taken from a spoken sample, the result of SIV processing.

**SIV extension**

The standard specification that will be created from this requirements document.

**SIV processing**

Any processing performed by an siv resource, for example, enrollment, adaptation, authentication, and buffering.

**SIV session**

A segment of the interaction with the user that is performed for the purpose of authenticating/enrolling the user. A single SIV session involves one or more types of SIV Processing with a single set of claims.

**spoof / spoofing**

Imitating the biometric of an authorized user (e.g., mimic, tape recorder)

**text dependent**

SIV technology (usually verification technology) that requires the voice input of one or more specific passwords or pass phrases (having been enrolled).

**text independent**

SIV technology that can operate on any freeform or structured spoken input

**text prompted**

(also called challenge-response) SIV technology (usually verification) that randomly selects words and/or phrases and prompts the speaker to repeat them.

**threshold**

The value above which the degree of similarity between two compared models is sufficiently high to return an "Accept" verification decision and below which the degree of similarity between two compared models is sufficiently low to constitute a "reject". Thresholds can often be adjusted at an administrative level to decrease the false match rate or to decrease the false non-match rate.

**turn**

A dialog with the user that consists of a single request and a single response. Same as Interaction Turn.

**utterance**

spoken input speech sample. May be real time streaming audio, a prerecorded file, or the result of buffering. In interactive systems a single utterance typically corresponds to a single interaction turn.

**verification**

The process of comparing an utterance against a single reference model based on a single claimed identity (e.g., user ID, account number). A verification result includes both a score and a decision. Same as - one-to-one.

**Voice model**

It is a system's representation of the an individual's voice and is constructed from data extracted from one or more voice samples provided by that individual. Also see reference model

**Voiceprint**

A synonym for voice model that has fallen out of favor because it improperly implies correspondence with standard fingerprint images.