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

Working Draft for Comment – 2008
http://www.voicexml.org/resources/biometrics.html

VoiceXML Forum Speaker Biometrics Committee

Editor: Valene Skerpac, iBiometrics, Inc.
Co-editors:
Chuck Johnson, AnyTransactions
Judith Markowitz, J. Markowitz Consultants
Ken Rehor, Cisco

Send comments to: valene@ibiometrics.com
Format: Name, Organization, Email,
General Document Comment,
Section Comment (please identify section).

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
July 31, 2006	Internal Working draft – Introduction section, Applications and Voice Engine Management, Outline of Best Practices section for review
May 21, 2007	Deleted terms in last section and made reference to the Glossary document and other SIV VoiceXML documents
August, 2008	Included updated Security, Architectures and Engine Management Section – Revamped to FAQ format

INTRODUCTION AND BEST PRACTICES

TABLE OF CONTENTS

1.0 Introduction

The growing scourge of fraud, identity theft, and cybercrime, demands increasingly powerful and effective authentication and identification. This has led to a rise in the use of biometrics. Among those biometrics are speaker identification and verification (SIV) which is being used as a primary source of authentication/identification and in multi-factor solutions in VoiceXML solutions.

There are several reasons for this growing market interest. Among the reasons for this popularity are the following:

- It operates with universal devices (telephones and microphones)
- It has been repeatedly identified by consumers as a highly-acceptable technology
- It is comparatively inexpensive
- It works well in interactive voice response (IVR) systems with speech recognition

These and other factors led the VoiceXML Forum to establish the Speaker Biometrics Committee (SBC) in 2005. Since then, the SBC has identified a number of projects that are designed to empower VoiceXML application developers interested in using SIV. Those projects include

- requirements for standardizing SIV as part of the VoiceXML language
- categorization of SIV applications
- liaison relationships with other standards bodies working on SIV - notably, the American National Standards Institute (ANSI) and the International Standards Organization (ISO)
- providing the speech-processing developer community with a guide for using SIV.

Documents produced by the first three projects are published on the SBC's page of the VoiceXML Forum's website (http://www.voiceXML.org/biometrics). The SBC has addressed the fourth objective in two ways: it has created a glossary of terms which appears on the website and it has created this introduction and best-practices document.

1.1 Goals of Document

SIV Introduction and Best Practices provides guidance in the form of Frequently Asked Questions (FAQs) for developers and organizations planning to implement Voice XML and other SIV applications – primarily speaker verification/authentication applications. The concept of "best practices" includes security and privacy considerations as well as application management and user considerations. Some of the broader topic areas, such as biometric security and privacy, have been addressed extensively in other publications. Rather than duplicate that existing work this document provides a foundation and then directs the reader to other resources.

In addition to broader topic publications, this document complements VoiceXML forum SIV publications (http://www.voicexml.org/resources/biometrics.html) including the SIV Glossary, Speaker Identification and Verification (SIV) Requirements for VoiceXML Applications SIV Applications, and Data Interchange documents under development. This document represents the first public draft of work directed at the fourth objective. We invite your comments and we look forward to working with you going forward.

1.2 Structure of the Document

Section 2 provides a comprehensive introduction to biometrics and SIV technology using systems models as an avenue to communicate a wide range of applicable information. Its flow takes the reader from the basics of biometrics, through the basics of SIV technology to its integration with Voice XML. This section is designed to serve as a handy reference.

The next 3 sections address specific aspects of SIV best practices. They utilize the method of Frequently Asked Questions (FAQs) to guide practitioners. The questions used to structure those sections reflect common issues that arise as an organization moves forward with SIV- especially speaker authentication/verification. That technique not only addresses important aspects of best practices it also provides important questions to ask vendors, integrators and services providers.

2. Overview of the Technology

How are biometrics used in information systems?

Biometrics statistically measure certain human anatomical, physiological and behavioral 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' as well as a method for assigning an identity to the voice of an unknown speaker That is, biometrics provides direct authentication of the user. All other user-authentication methods are indirect.

Indirect methods authenticate the user through one of the following:

- Something they know (e.g., PIN, password, mother's maiden name, account number)
- Something they have (e.g., key, token, badge)
- Their location (e.g., GPS)
- Their behavioral patterns (e.g., behavioral profile).

Characteristics used commercially as biometric representations of the user include voice, fingerprints, hand geometry, palm, iris patterns, retinal patterns, facial image, and sign/signature verification, and keystroke dynamics. 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

How do systems process biometrics?

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 basic steps are involved in biometric verification and identification:

- Input of the biometric
- Quality analysis and potential re-capture of the biometric input
- Creation of digital representation of the captured biometric
- Comparison of that digital representation with previously enrolled representation(s) to determine if a match exists
- Scoring of the comparison
- Generation of a decision based on that score.

There are three biometric processes associated with biometric authentication:

> 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.

Verification: Verification confirms that the user is who she/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.

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 form 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. This is a one-way process. That is the biometric digital representation cannot not be converted back to the raw data.

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.

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. 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:

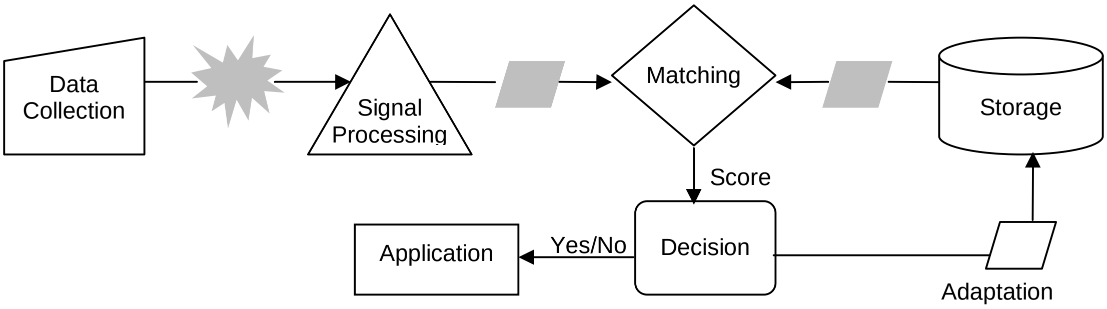


Figure 1 *Generic Biometric Model*

2.2. Speaker Identification and Verification Technology

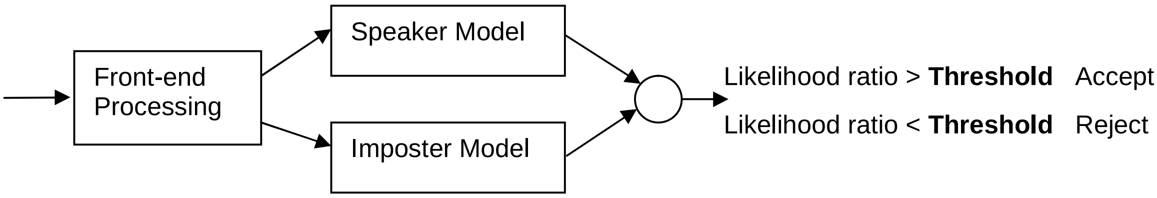
How can SIV biometrics technology uniquely verify or identify someone's voice?

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. *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 applications are more flexible.

There are various technical approaches to SIV. One predominant method employed in today's SIV products is the use of likelihood ratios as described in 'An Overview of Automatic Speaker Recognition Technology' by Douglas A. Reynolds of MIT Lincoln Laboratory (Ref). 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 user's 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 *Speaker Verification Technology Model*


Front-end Processing
Speaker Model
Imposter Model
Likelihood ratio > **Threshold** Accept
Likelihood ratio < **Threshold** Reject

How Does SIV technology model map to the generic biometric model?

The speaker verification technology model maps to the generic biometric model to create a generic SIV model, as shown in **Figure 3.** This mapping is helpful to organizations that implement and manage SIV systems so that they can utilize biometrics standards such as ISO 19092 which is a standard for biometric security management,

The following figure illustrates a verification (only) process using the major components of the generic SIV system model and correlates them to a traditional legacy Packet Switched Telephone Network (PSTN) environment as an example of where the components run:

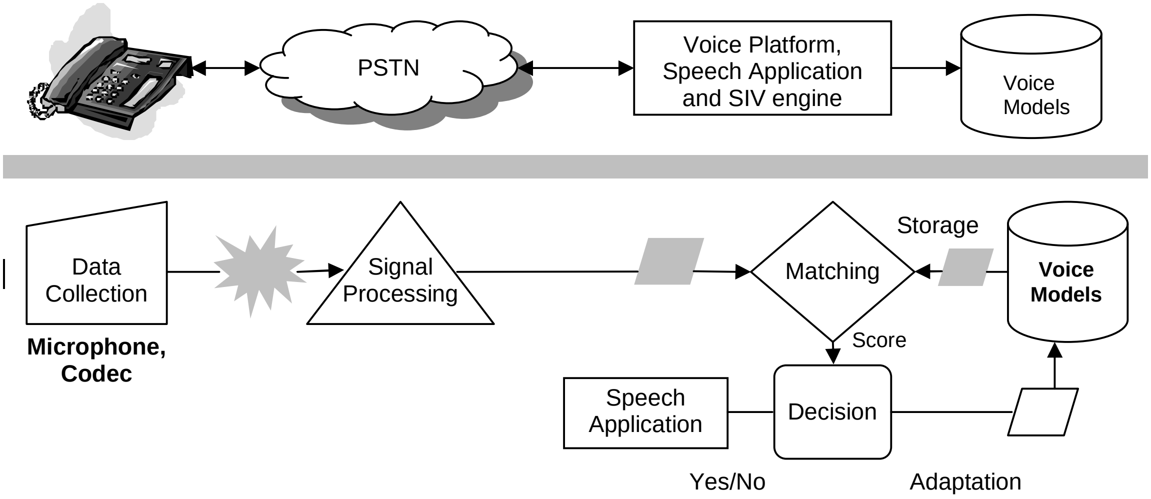


Figure 3 *Generic SIV Model mapped to legacy PSTN*

Major Components of a generalized **SIV model** 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.

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[1] that is representative of the speaker. There are a number of modeling techniques used to create an appropriate voice model[2]. 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 technology section, the score matching process leads to a **Likelihood Ratio, which is compared to a Threshold to decide to accept or rejec**t. 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 devices pr 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.

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

How does the preceding SIV generic model relate to the developing SIV specifications for the Internet, specifically those from the VoiceXML Forum and WC3 standards organizations?

[1] Other terms Speaker Model or Voice Print are sometimes used to refer to the voice model.

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

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.

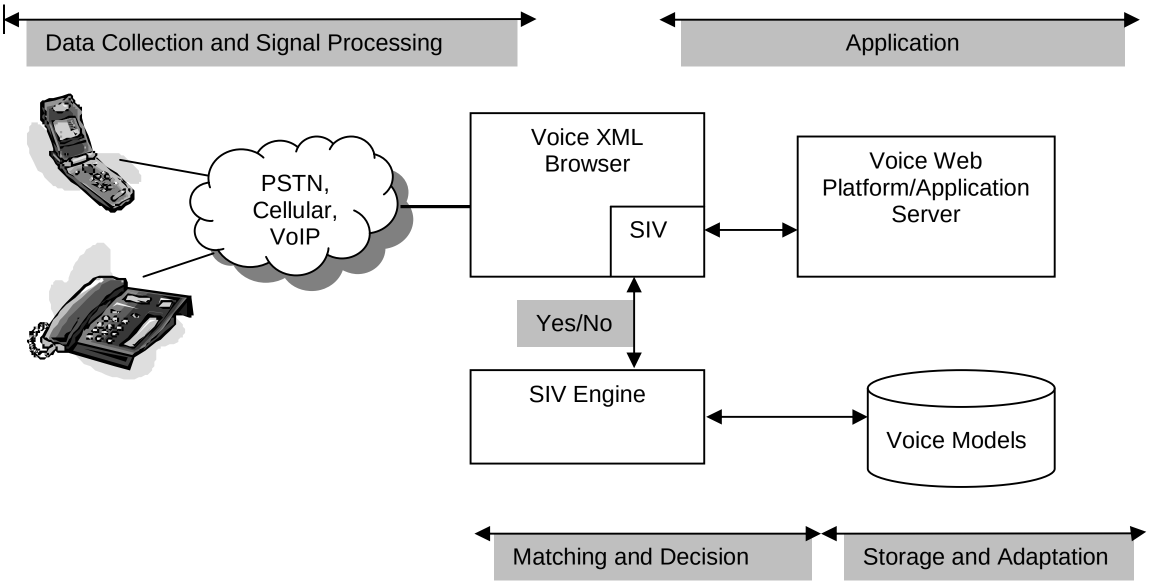


Figure 4 *VoiceXML SIV Architecture mapped to SIV Model Components*

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.

3. Applications

3.1 Feasibility Assessment

Can SIV be used with my current or planned VoiceXML applications and what are the current limits of the technology?

SIV feasibility is an important aspect of the initial planning and discovery phase of a speech application. SIV can be used in any application that benefits from greater knowledge about the identity of a speaker or group of speakers. It should be considered as one valuable piece of identity information amongst other identifiers; it should not be used as the only identifier of the user. Current limitations of SIV technology and the realities of 'less than perfect' real world computing environments preclude 100% SIV accuracy.

Although there are many aspects of user authentication that cross application boundaries, the feasibility of an SIV (or other biometric) application often has unique challenges and should be assessed on an application basis. These challenges are within known categories, such as backend integration, voice user-interface (VUI), and even the nature of the SIV dialog.

3.2 Project Life Cycle Approach

Can I use my organization's established approach to managing speech application development and deployment when I add a new SIV speech application or add SIV to an existing speech application?

Yes, the established speech project lifecycle methodology used to develop and deploy successful speech applications can include SIV tasks. The speech project lifecycle recognizes five phases which are Planning and Discovery, Design, Development, Deployment and Tuning and Maintenance (ref.).

The planning and discovery phase of a project collects information used as the basis for a design requirements specification document which includes a vision that defines the objective and tradeoffs of the project. This phase includes activities such as business case, application and design specifications, feasibility, costs, stakeholders, timeframes, policy and regulatory security and privacy requirements. It can also include the collection of application-specific data and documentation such as existing scripts logs, system models, projections, service standards and audits. During this phase, the feasibility of SIV needs to be determined as well as its effect on the business case, application, design and timeframes.

The output of the design phase is an approved design requirements specification which insures that priorities and expectations are set appropriately. Tasks typically performed in the design phase are the design of SIV user interface, interaction modules with SIV engine APIs or web service calls, design and documentation of the entire dialogue flow and call flow design. Agreement is reached on feature priorities, timeframes, how the product will be built and who will build it, product architecture, risks of the product, and milestones and deliverables during the project.

Organizations may require a proof-of-concept system and pilot during the design and development phases of a new application to do the following:

- demonstrate how enrollment and verification prompts sound within the voice interface
- assess usability of SIV for their application focusing on the Voice User Interface (VUI)
- assess SIV performance (see engine management section) for their application

The development phase is tasked with building the SIV speech application as defined in the previous phases. Common efforts in this phase are to develop detailed specifications and test plans, develop dialogues and unit tests, select voice talent, code and unit test applications, write and test grammars, control and track audio engineering, integrate code, grammar and audio, perform usability and acceptance testing, debug, tune and iterate.

The deployment phase commences when the operations and support groups are officially responsible for ongoing maintenance and support. It is important that controlled design and development practices are followed to avoid 'Function Creep' that can jeopardize security and privacy compliance.

The tuning and maintenance phase of a speech application includes the capture of actual data from callers and subsequent use to refine or tune the grammars, engine parameters, and other aspects of the dialogue. The requirement to tune speech applications after deployment is due to the fact that only actual caller data is truly representative of the target population. This phase' activities can include monitoring of logged calls and reports, identification of specific problem areas such as a particular grammar or prompt, bug fixes and synchronized system updates. During speech application tuning, performance logs should be protected because they often contain sensitive information.

SIV performance and tuning can be done in the earlier phases or is sometimes done after deployment. The tuning and maintenance phase that focuses on the SIV performance is discussed in the engine management section of the document.

3.3 User Interface Design

When designing an SIV User Interface, what questions/issues should be considered?

- How technologically savvy are my users?
- How often will people use the system?
- What kind of help and assistance might the users need?
- What are the methods and procedures if voice authentication fails?
- What kinds of telephones will they be using and will they switch among them?
- Are there age considerations?

3.4. Other Identifiers used

How can my SIV application use other identification factors?

SIV applications today incorporate voice authentication as a second identification factor. Non-biometric factors to consider during authentication include a PIN, password, caller-ID, and answers to knowledge questions. Each organization should consider the confidence they have in each factor and the risk associated with a transaction or account access. This assessment is discussed in the security section of the document.

4.0. Voice Engine(s) Management

This section addresses the issues and frequently-asked questions regarding SIV engine management.

What types of verification are supported by the SIV engine?

Depending on the vendor, the SIV engine will support one or more of the following types of speaker verification:

- Text Independent

  Text independent is an SIV technology that operates on any freeform or structured spoken input. One simple example of text independent verification is when a designated user has been enrolled via speech data collected through a series of free form prompts.
- Text Dependant

  Text dependent SIV technology (usually verification technology) requires the voice input of one or more specific pass-phrases (having been enrolled). One simple example is when a user has been enrolled via voice data collected through a series of prompts to speak a designated pass-phrase (such as an account number, 123456789).
- Text Prompted

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

When selecting a type of speaker verification, the following questions/issues should be considered:

- What types of speaker verification are the 'best' for your application?
- What types of speaker verification are supported [by the vendor]?
- What is the minimum amount of speech/audio needed for verification?
- What are the 'voice model creation requirements' for each type of supported verification?

What data storage infrastructures are needed for SIV?

To store voice models (voiceprints) and 'knowledge verification' information, an SIV system requires a data storage infrastructure. For voice model storage, the following questions/issues should be considered:

- What are the supported storage infrastructures?
- Is a DBMS system required for storage?

  If so, what databases are supported? What database interfaces (i.e. ODBC) are supported? How is the database configured and administrated (for storage and access)?
- How is a voice model accessed?
- How is a voice model secured (within/by the storage infrastructure)?

What are verification scores and decision results and what do they mean?

In ASR (Automatic Speech Recognition), the recognition engines returns one/more recognition result / hypothesis and a confidence score for each hypothesis. The speech application can evaluate the confidence score and compare it to a pass/fail threshold to determine whether to accept or reject the recognition result.

In Speaker Verification, the SIV engines (depending on the vendor) can be configured to return a numeric verification score, or a verification decision, or both.

- Raw score

  A numerical representation of the degree of similarity between data processed from a voice sample and a reference model. The specific method, by which a score is generated, as well as the probability of its correctly indicating a match / non-match, is generally propriety to each engine vendor.
- Normalized score **-** The normalized raw score
- Pass/fail or match/mismatch decision - Binary decision to accept or reject
- Match/mismatch/inconclusive decision

When interpreting verification results, the following questions/issues should be considered:

- Does the engine return raw/normalized scores? If so, what do the raw and/or normalized scores mean (with respect to the verification decision)?
- Does the engine return a confidence score? If so, what does the confidence score mean? How should the confidence scores be used to make/support a verification decision?
- Are there ‘degrees’ of pass/fail or match/mismatch? If so, how are they generated? How should they be interpreted?
- In addition to scores and/or decisions what other results are returned (i.e. error, exception, partial results, logging)?

SIV Decision management for the undecided or grey area

In more advanced SIV decision applications, a score can be combined with a confidence value based on business risk and policy to make a yes or no decision. The confidence measure is usually one of a number of levels, or a normalized score where the higher the number, the higher the confidence

In practice, a threshold provided by the SIV engine vendor is often used whereby the score must not exceed the prescribed threshold. During tuning, FA (False Acceptance rate) and FR (False Rejection rate) rates are measured given various conditions such as the amount of speech, channel noise, device and speech quality. It is the tradeoff of these FA and FR errors under an assumed set of conditions which is used to develop the threshold. Some vendors provide a range of scores that fall below the threshold often referred to as the “grey area” or “undecided” that could be considered valid scores depending on conditions.

At the system and application level (or even transaction), an organization may want to consider factors that make up a confidence value based on business risk and policy to assess fully the score results. For example, if the telephony application has positively matched several other identity factors of the caller, the organization may have a policy to pass scores that fall within the top 20% of the grey area for low to medium risk transactions.

The following figure illustrates a more advanced authentication decision process:

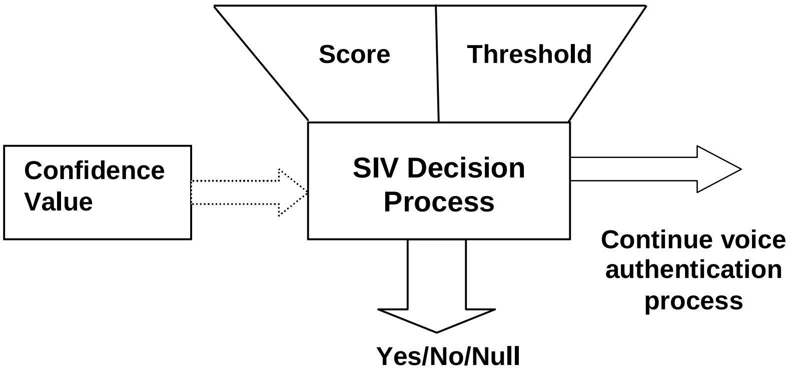


Figure 5 Advanced SIV Decision Process

What are verification threshold(s) or security levels and how are they set?

Setting the SIV operating point/threshold entails balancing the false match rate (aka false accept rate) with the false non-match rate (aka false reject rate).

- False Match Rate - Same as False Acceptance Rate (FAR)

  In a One-to-One matching verification system, the FAR is the probability that a system will falsely verify an imposter as a legitimate enrollee. In a One-to-Many matching identification system, FAR is the probability that a system will incorrectly identify an individual.
- False Non Match Rate - Same as False Rejection rate (FRR)

  In a One-to-One matching verification system, FRR is the probability that a system will fail to verify the identity of a legitimate enrollee. In a One-to-Many matching identification system, FRR is the probability that a system will fail to identify a legitimate enrollee.

The setting of the operating point/operating threshold will determine the false match/false accept and false non-match/false reject rates of the SIV engine. Some SIV engines allow/support the setting of the match and/or non-match rates. Other engines support the setting of 'verification thresholds' that in turn establish the match and non-match rates. Some SIV engines support the concept of Verification Security Levels. Typically the security level corresponds to preset thresholds or preset ranges of false match/false accept rates.

When setting the threshold, the following questions/issues should be considered:

- Does the engine have default operating point/threshold or security level? Is so, what is it and what does it mean?
- Can the operating point/threshold(s) of the engine be changed/modified? If so, how do you establish the operating point of the SIV engine?
- Does the engine support the setting of security levels? If so, what are the levels and what to the mean?
- Does the engine support the setting of false match and/or non-false match rates? If so, how are the rates established and set?
- Can the security level/threshold be set at run-time for specific user, groups and/or individual applications?

What factors affect the performance of deployed systems?

The verification performance of deployed systems can be affected by a variety of factors. When deploying SIV applications, the following questions/issues should be considered:

- How is the verification performance affected by noise?
- What is the impact of cross channel affects?
- What is the verification performance when using mobile devices?

What is voice model (voiceprint) adaptation?

Adaptation is the process of updating or refreshing a reference voice model. 'Supervised adaptation' is usually invoked by the application based on application-specific criteria. 'Unsupervised adaptation' is typically performed automatically by the engine if it determines that the user is the true speaker. Adaptation will ensure that the quality of the voice model, and therefore system performance, will improve over time.

When considering voice model adaptation, the following questions/issues should be considered:

- Does the engine support voice model adaptation? Is so, what types are supported?
- Is adaptation enabled by default or does it have to be enabled ('turned on')?
- If supervised adaptation is supported, how does the application 'decide' when to adapt the voice model?
- If unsupervised adaption is supported, how is the engine configured to adapt the voice model?

<u>Advanced Topics to be further</u>

1) SIV Accuracy Tuning
Note: Will not be included in Draft for SpeechTek2008

2) SIV Performance Evaluation
Note: Will not be included in Draft for SpeechTek2008

3) SIV and ASR Management
Note: Will not be included in Draft for SpeechTek2008

5.0 Security

This section addresses some of the frequently-asked questions regarding security and SIV. Since security is a broad, yet sensitive, topic we welcome input and participation from security specialists and security organizations.

How do I determine how much security my application actually needs?

Your corporate security policies and procedures will provide a great deal of guidance in this regard. In addition, there are a number of standards and guidelines that provide assistance with regard to this decision. ISO 19092 (REF) provides guidance to an organization to perform an SIV risk assessment which dictates how much security is required. Among other guidelines is the United States' Office of Management and Budget (OMB) Memorandum 04-04 *E-Authentication Guidance for Federal Agencies*. The memorandum is an example of federated authentication. It defines four "assurance" levels for authentication for Federal Government applications. The term "assurance" refers to the level of confidence that the person presenting her/himself to a system is who they claim to be. The level refers to the degree of assurance needed for that application.

Determination of assurance levels is accomplished through a risk assessment for the transaction that identifies risks and the likelihood of those risks occurring. The OMB provides the following table as a summary of those risks (left column) and the likelihood the risk will occur (rows containing "low – mod - high"). The assigned assurance level is in the second row (showing 1-4). The memorandum also provides a great deal of supporting information to guide your determine of assurance level.

Table 1 – Maximum Potential Impacts for Each Assurance Level

Assurance Level Impact Profiles				
Potential Impact Categories for Authentication Errors	**1**	**2**	**3**	**4**
Inconvenience, distress or damage to standing or reputation	Low	Mod	Mod	High
Financial loss or agency liability	Low	Mod	Mod	High
Harm to agency programs or public interests	N/A	Low	Mod	High
Unauthorized release of sensitive information	N/A	Low	Mod	High
Personal Safety	N/A	N/A	Low	Mod or High
Civil or criminal violations	N/A	Low	Mod	High

The OMB's hierarchy has been adopted by a number of biometric standards and guidelines, including National Institute of Standards in Technology (NIST) SP 800-63 *Electronic Authentication Guideline Study Report on Biometrics in E-Authentication* (SP 800-63), American National Standards Institute/InterNational Committee for Information Technology Standards (ANSI/INCITS), and the International Standards Organization (ISO) *Financial services — Biometrics — Security framework* (ISO 19092). The ANSI/INCITS study report and ISO 19092 apply the OMB's hierarchy to biometrics.

SIV is appropriate for all four assurance levels but levels 3 and 4 require multi-factor authentication.

NIST just released the working draft of a guidance that will also be of use. Its focus is on security for cell phones and PDAs. *Guidelines on Cell Phone and PDA Security* (SP 800-124). The draft was released for comment in July, 2008. It is part of a series of publications on computer security issues.

Why do I need SIV? Aren't PINs and passwords enough?

The stunningly high entropy levels tied to long, arcane passwords that are changed frequently do not translate into greater real-world security. One of the most compelling reports documenting the failure of PINs and passwords is by Trusted Strategies. In 2006, they released Cybercrime Study entitled *Network Attacks: Analysis of Department of Justice Prosecutions 1999 – 2006*. Among their key findings were

- Organizations suffered the greatest financial loss and damage, more than $1.5 million per occurrence, when attackers used stolen IDs and passwords;
- Losses from stolen IDs and passwords far exceeded damages from worms, viruses, and other attack methods not utilizing logon accounts;
- 84% of the attacks could have been prevented if the identity of the computers connecting were checked in addition to user IDs and passwords.

Even though the reference is to the use of PINs and passwords to access data directly via computer the findings can easily be extrapolated to PINs and passwords keyed into a telephone or spoken to a speech-recognition system.

Unless the assurance level of the task/transaction is at Level 1 (see question 1, above) PINs and passwords are not to be trusted. When used by themselves they are even suspect for Level 1 operations because they can be stolen, shared, lost, or the captured using keystroke-capture technology.

Certainly, SIV is not the only method for supplementing or replacing PINs and passwords but it is a viable and more secure replacement or partner for them.

If SIV can't give me 100% accuracy why should I use it?

Any experienced security professional will admit that every form of security known has vulnerabilities. Nothing is 100%. Furthermore, test results from laboratories only indicate that a technology works under ideal or controlled conditions. Even non-laboratory performance studies need to be carefully evaluated to determine their utility for your application, environment, and user population.

Steps to help you determine whether SIV is suitable for your application include the following:

- Apply the security policies and procedures of your organization;
- Determine the security/authentication assurance level needed by your application (see question 1, above). ;
- Identify the alternate security options for your application given its level, access methods (e.g., telephone), and the population of users. They may include multi-factor authentication or placing limits what a person can do (e.g., withdrawal of funds from some ATMs is limited to $300);

- Identify the known vulnerabilities of each alternative (for biometrics and SIV the INCITS *Study Report on Biometrics in E-Authentication* and ISO 19092 are useful); and

- Perform feasibility, human factors, and user acceptance analyses.

These activities will also help you determine reasonable error rates for your application, potential sources of vulnerability, and whether/how a multi-factor solution might be useful.

Also consult other sections of this document, especially Voice Engine Management, which will help you determine, for example, whether the out-of-the-box settings of an SIV product are right for your application.

How do I determine whether my SIV application is actually working the way it is supposed to work?

The best way to ensure that your application is working properly is to perform periodic audits as part of the application's life-cycle management. Those audits should look at each component of the system (data collection, matching, storage, data access, etc.) for the system's major functions: enrollment, re-enrollment, verification, deletion, and identification (if used). Since the SIV system is part of your organization's larger security structure the audits should follow your organization's audit schedules, policies, and procedures.

For more information consult ISO 19092 and the Voice Engine Management section of this document.

How do I make my application both secure and easy-to-use?

Too often, security and convenience are seen as antagonistic concepts. One of the great benefits of SIV is that it demonstrates that it is possible to provide both. In fact, SIV is sometimes selected more for its ability to enhance convenience than because of the security it offers.

Creating a system that is both convenient and sufficiently secure involves a balancing act. Memorandum 04-04 *E-Authentication Guidance for Federal Agencies* of the US Office of Management and Budget, ISO 19092, and INCITS *Study Report on Biometrics in E-Authentication* provide important guidance related to the security side of the equation.

To ensure that the voice user-interface (VUI) of an SIV system is both convenient as well as secure, the developer must be skilled in human-factors design and needs to understand security vulnerabilities. Keeping in mind the security level of your application (see Question 1, above) is useful for retaining the security component. It is not reasonable to demand that someone wanting to get the bank account balance to repeat three strings of 12 digits or answer 5 personal questions. Conversely, it is not secure to permit someone requesting funds transfer to easily back down to a simple password. For example, it isn't reasonable to demand that someone wanting to get the bank account balance to repeat three strings of twelve digits or answer five personal questions. Conversely, permitting someone requesting a funds transfer to easily back down to a simple password will not provide sufficient protection.

Do I need to be concerned about privacy?

There are many definitions for the term "privacy." The one that applies to SIV and other data systems is protection of personal data. There are several reasons you need to be concerned about this kind of privacy:

1. SIV applications often process personal data, such as account numbers and personal/employee IDs;

2. Biometric data, including SIV voice models, are, themselves, considered to be personal data;

3. In some countries and localities privacy protection has the weight of law. The first regulation to be put into place was the European Union's 1998 *Data Protection Directive*. Since then, other nations, including Australia, Canada, Israel, and Japan have published their own privacy regulations and appointed privacy commissioners. Two of these regulations appear in the References;

4. Securing personal data and instituting best-practices for protecting privacy is nothing more than smart business. It's no fun dealing with angry customers and/or employees, determining which data have been lost or compromised, making restitution, or suffering bad publicity.

There are several principles shared by many of the privacy regulations. They are:

1. *Purpose*: Personal data must be collected and possessed for a clearly-defined and legitimate purpose and kept no longer than necessary to fulfill the stated purpose.

2. *Limitation of Use*: Your organization must not use any personal data for purpose other than the stated, primary purpose for which the data were collected. This is to prevent function creep. There are a few exceptions, such as when the individual agrees to the new use of the personal data;

3. *Consent*: The individual providing the personal data must be informed why the personal data are being taken and how the data will be used (see *Purpose*). That individual must be providing the data willingly;

4. *Data Protection*: All reasonable steps must be taken to secure personal data. Best practices for accomplishing this include data encryption, access control, and separating personal data from other information. ISO 19092 provides a great deal of useful guidance on this topic;

5. *Disclosure and data transfer*: The individual must be informed of and approve any plan to disclose or share personal data with outside individuals or groups. This includes sharing the data with other groups, departments, or agencies within your organization. Disclosure includes the publication of personal information through any medium. It often also includes accidental disclosure and theft. Some laws prohibit sharing data with any country or entity lacking "adequate level of protection."

6. *Data Quality*: Data must be accurate and up-to-date. Best practice includes adaptive updating of SIV voice models.

7. Individual Redress: An individual who provides personal data to your organization must have the right to

- access her/his personal data;
- correct or block inaccuracies;
- object to the use of those data.

What do I do if the SIV voice models are stolen or tampered with?

You should utilize the audit and security procedures of your organization to determine what occurred and which data were affected. Your response to such violations should be based on those procedures and data protection principles, such as principle 5 in the answer to the previous question.

6. References

6.1 VoiceXML Documents (http://www.voicexml.org/resources/biometrics.html)

[Apps] "Speaker Verification and Identification Applications" by VoiceXML Forum Speaker Biometrics Committee.

[Architecture] "Speaker Verification and Identification Architectures and Data Structures" by VoiceXML Forum Speaker Biometrics Committee (in preparation)

[DEFF] "Data Exchange File Format for SIV" by VoiceXML Forum Speaker Biometrics Committee

[Glossary] "Speaker Identification and Verification (SIV) Glossary" by VoiceXML Forum Speaker Biometrics Committee.

[Requirements] "Speaker Identification and Verification (SIV) Requirements for VoiceXML applications" by VoiceXML Forum Speaker Biometrics Committee

6.2 External References

An Overview of Automatic Speaker Recognition Technology, Douglas A. Reynolds, MIT Lincoln Laboratory, MA, USA, **www.ll.mit.edu/IST/pubs/020513_Reynolds.pdf**, 2002.

Speaker Recognition: A Tutorial, Joseph P. Campbell, Jr., Proceedings of the IEEE, Vol. 85, No. 9, September, 1997

The Speech Project Lifecycle, Microsoft MSDN Library, msdn.microsoft.com/library/default.asp?url=/library/en-us/SASDK_UserManual/html/UM_design_LifeCycle.asp.

Biometrics Institute 2006 *Privacy Code*. Crows Nest, Australia.

European Commission 1995 (Directive 95/46/EC) *On the Protection of Individuals with Regard to the Processing of Personal Data and on the Free movement of Such Data*. Brussels, Belgium.

InterNational Committee for Information Technology Standards 2007 *Study Report on Biometrics in E-Authentication*. Washington, DC.

International Standards Organization 2008 (ISO 19092) *Financial services — Biometrics — Security framework.* Geneva, Switzerland.

National Institute of Standards in Technology 2004 (NIST SP 800-63) *Electronic Authentication Guideline.* Gaithersburg, MD.

National Institute of Standards in Technology 2008 (NIST SP 800-124-draft) *Guidelines on Cell Phone and PDA Security*. Gaithersburg, MD.

Office of Management and Budget of the United States 2003 (Memorandum 04-04) *E-Authentication Guidance for Federal Agencies.* Washington, DC.

Trusted Strategies 2006 *Network Attacks: Analysis of Department of Justice Prosecutions 1999 – 2006.* Pleasanton, CA.