Best Practices

This document contains adapted portions of a document developed by the VoiceXML Forum Speaker Biometrics Committee (2008). The document uses a FAQ approach.

I. Application Development

A Feasibility Assessment

Can Speaker identification and Verification (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.

B 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 Section II Voice Engine(s) Management) 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 Section II Voice Engine(s) Management.

C. 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?

D. Other Identifiers used

How can my SIV application use other identification factors?

SIV applications today incorporate voice authentication as a second identification factor.

Nonbiometric 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.

II. 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 SV engine will support one or more of the following types of input for 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. The power of this approach is that it can be used in the background while a user is speaking to a human or a speech-enabled, IVR system.

Text Dependant: Text dependent 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 (also called “challenge-response”) prompts the user to repeat one or more words, numbers, and/or phrases (e.g., “Say *1 2 5 3 7*”, “Say *black coffee*,” “Say *34 96*”). The items are generally randomly selected from among the words, numbers, and/or phrases that the user has already enrolled. The greatest value of this approach is to defeat tape recordings (called “tape attacks” or “spoofing”). Some engines go further by using the enrolled material to generate novel sequences – sequences the user has never spoken to the system. If, for example, the user enrolled the digits 0 1 2 3 4 5 6 7 8 9 such an engine could generate a prompt asking for any re-ordering of those digits.

Speaker identification engines are fundamentally text-independent but can be constrained to operate in either of the other modes.

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 a DBMS system is required for storage the following questions apply:

- 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 Automatic Speech Recognition (ASR) 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. Here are the things that are typically returned:

Raw score: A numeric 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: A binary decision to accept or reject the claim of identity in SV applications.

Inconclusive decision: When the matching score is not clearly accept or reject. This is typically associated with applications that employ “gray areas.” Gray areas are discussed in the next section.

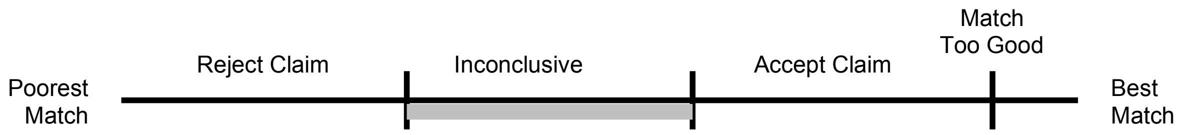
When interpreting verification results, the following questions/issues should be considered:

- 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 gray 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.

Such applications utilize two or more thresholds to make various decisions, as shown below



There are three thresholds in the above example. The thresholds demarcating the “inconclusive” decision represent the gray area. The scores are too poor to confidently accept the claim and too good to confidently reject the claim. The “Match Too Good” decision is one way SIV systems detect possible tape attacks.

During tuning, False Acceptance (FA) and False Rejection (FR) 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 “gray 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 gray area for low to medium risk transactions.

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

Setting the SIV operating point/threshold entails balancing the FAR with the FRR

False Acceptance Rate (FAR): In most SV applications, the FAR is the probability that a system will falsely verify the identity claim of an imposter. In most SI applications, FAR is the probability that a system will incorrectly identify an individual.

False Rejection Rate (FRR): In most SV applications FRR is the probability that a system will fail to verify the identity claim of a legitimate enrollee. In most SI applications 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?
  - If so, what is it?
  - 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: Supervised adaptatation is usually invoked by the application based on application-specific criteria.

Unsupervised adaptation: 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 adaptation is supported, how is the engine configured to adapt the voice
- model?

III. 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).

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 memorandum also provides a great deal of supporting information to guide your determine of assurance level.

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.

In July, 2008 The US National Institutes of Standards in Technology (NIST) 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). 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 Section II Voice Engine(s) 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 Section II Voice Engine(s) Management and International Standards Organization document ISO 19092.

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 EAuthentication* 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.

IV Privacy

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.

V. References

The following are cited in the text or useful references

Microsoft *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) *EAuthentication Guidance for Federal Agencies.* Washington, DC.

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