



Speaker Identification and Verification (SIV) Glossary

Draft - May 1, 2007

Editor: Valene Skerpac, iBiometrics, Inc.

VoiceXML Forum Speaker Biometrics Committee
Judith Markowitz and Ken Rehor, Co-Chairs
<http://www.voicexml.org/biometrics>

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
August 28, 2006	Internal Working draft – Consolidation of SIV terms in one master document
May 01, 2007	Public draft

1. Document Overview

This document establishes a common list of terms for speech and speaker recognition technologies.

2. Terminology

This section includes an enumeration of terms and abbreviations related to biometrics technologies and SIV technology.

accept

Speaker verification and identification are comprised of *matching* followed by a *decision* to accept or reject. The speaker verification process can decide to 'accept' an individual as a result of *one-to-one matching*. The speaker identification process can decide to 'accept' an individual as a result of *one-to-many matching*.

adaptation

Adaptation is the process of updating or refreshing a reference model. 'Supervised adaptation' is usually invoked by the application based on application-specific criteria. 'Unsupervised adaptation' is typically performed automatically by the engine on a pre-determined basis.

ASR

Acronym for automated speech recognition. See speech recognition.

attempt

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

attribute

Attribute is an XML term that provides additional information associated with each XML tag or element. For example, in Voice XML 2, the attribute of a <prompt> element might be the particular character string that the prompt will present to a user. The attribute of a <go to> is the address of the location that the control should be passed to.

authentication

Authentication is the process of verifying a user, device, or other entity often as a prerequisite to allowing accesses to resources in a system. Within SIV, the term is synonymous with verification since a *claim of identity* is verified (using speaker verification).

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

Buffering is 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 name or index of a reference model (out of the set of reference models) that the SIV system detected as the speaker (that spoke the input voice sample).

capture

The acquisition of a spoken sample.

challenge-response

Synonym for *text prompted*.

claim of identity

The act of claiming an identity for the purposes of verification. See *identity claim*.

claimant

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

closed set identification

An identification process is said to be Closed-Set if the test speaker is known to exist in the trained database. In this case, the identification engine is only required to return the identity of the person in its database whose model most closely resembles the test sample. N.B., The closed-set identification process is not expected to be able to reject the test speaker.

co-located resources

SIV and ASR are combined in one engine. Typically they share some processing phases like endpointing, feature extraction etc. and have a set of shared return results like non-match/non-input.

confidentiality

Confidentiality is one of the basic security requirements. It is the protection of data, including SIV data, from unauthorized access and inadvertent disclosure.

decision

The process of deciding a match / non-match based on a *decision policy*.

decision policy

The criteria through which an SIV system bases its match / non-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

decryption

The process of transforming encrypted data back into readable data. Also referred to as decipher or unscrambling.

dialog

A dialog is the interaction between a user and computer during the length of a session.

digital signature

The use of public-key algorithms to simulate the security properties of a signature in digital, rather than written, form. Digital signature schemes normally give two algorithms, one for signing which involves the user's secret or private key, and one for validating signatures which involves the user's public key. The output of the signature process is called the "digital signature." Some digital signature schemes can ensure non-repudiation as well as detect any changes to the message content.

encryption

The process of transforming readable data into unintelligible form to hide its substance. Also referred to as encipher or scrambling .

enrollee

The individual enrolled or designated to be enrolled in the SIV system.

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 voice sample, 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 matching 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 (FAR).

false non-match rate

In a One-to-One matching 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 (FRR).

group authentication

A speaker biometric modality where an identity claim is made which pertains to membership in a group. A biometric system uses a group of records in a database for reference matching against a submitted biometric sample (processed voice sample) and if matched returns a positive verification. Also referred to as multi-verification. The term 'small set' or 'small group' identification has been used improperly as a synonym for group authentication in the past.

identification

Speaker identification is a biometric modality that uses an individual's speech for identification **without** a *claimed identity* (see *identity claim*). Identification is a task where the biometric system searches a database for a reference model *matching* a submitted biometric sample (processed voice sample) and if found, returns a corresponding identity.

A processed voice sample is collected and compared to all the reference models in a database. See Open Set and Closed Set definitions for more specifics. Identification is also referred to as one-to-many matching.

identity claim

Speaker verification associates an identity claim with the name or index of an individual reference model/enrollee or a group of related enrollees (group identity claim). An identity claim can be explicit or implicit from the user's perspective. An example of an individual *identity claim* is when a customer speaks their account number (1234578) which is associated with a reference model in the database. An second example of an *identity claim* is when an employee speaks their department name (Finance) which is associated with a group of reference models in the database.

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. Opposite of true speaker.

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.

initial enrollment

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

integrity

One of the basic security requirements. Integrity is the protection of SIV data from undetectable modification and substitution. Often implemented via Message Authentication Code (MAC) or digital signature techniques.

match(ing)

Matching compares processed data from an input voice sample against a previously stored reference model and scoring the degree of similarity or correlation between the two as being a match. Authentication comprises of *matching* followed by a *decision* to accept or reject

message authentication code (MAC)

A one-way algorithm that uses a secret key to authenticate a message. It allows for validation using the secret key to detect any changes to the message content. Also referred to as a data authentication code (DAC) or message integrity code (MIC).

modality

mode of operation

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"

Non-match/mismatch

Antonym to match. Not to be confused with speech recognition no-match.

non-repudiation

Protection of SIV data from renunciation and denial of issuance. Often implemented via a digital signature technique.

one-to-many matching

See Identification.

one-to-one matching

See Verification.

open set identification

An identification process is said to be Open-Set if the test speaker does not necessarily have any representative reference model in the database of the recognition engine. In this case, the speaker recognizer may either return an identity for the test speaker or may reject the speaker as not-present in its database. Some implementations of the open-set identification are done by doing a closed-set identification call followed by a consequent speaker verification call to the same or possibly different recognition engines.

parameter

A parameter is a value that you pass to a VoiceXML subdialog or object. For example, in Voice XML 2, the <param> element is used to specify values that are passed to subdialogs or objects. It is modeled on the [\[HTML\]](#) <PARAM> element.

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.

prompt

A request for action, for example audio response.

property

Properties are characteristics of the platform environment that can be overridden by the application developer. For example, in VoiceXML 2, properties are used to set values that affect platform behavior, such as the timeouts, caching policy, etc.

raw voice data

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

re-enrollment

Re-enrollment is the process of *enrolling* an individual with new voice data to create a new reference model replacing the old reference model.

reference model

The reference model is the data that represents the voice measurement of an enrollee. It is based on data extracted and processed from one or more voice samples provided by

that individual and is typically stored and used by an SIV engine for comparison against subsequent submitted voice samples. Reference model is the preferred term but it can also be referred to as voice model, template or voiceprint.

reject

Speaker verification and identification comprise of *matching* followed by a *decision* to accept or reject. The speaker verification process can decide to 'reject' an individual as a result of *one-to-one matching*. The speaker identification process can decide to 'reject' an individual as a result of *one-to-many matching*.

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.

response

A user reaction to a prompt; action may be an audio response.

rollback

A mechanism to "undo" or roll back the processing performed on the last turn (and only the last turn) in the enrollment and verification/identification session.

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.

scoring

The process of creating a score. See 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 the SIV requirements document.

SIV processing

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

SIV session

An SIV session is a segment of interaction between a user and a computer that performs speaker identification or verification or enrollment of the user. An SIV session usually consists of multiple turns. A single SIV session can involve one or more types of SIV processing depending on the particular 'use case' as specified in the SIV requirements document. An SIV session is distinct from a Voice XML session. Adaptation is considered an attribute of the SIV session

speaker classification

The process of categorizing different speakers based on shared attributes such as age, gender, accent, etc.

speaker recognition

A biometric modality that uses an individual's speech, a feature influenced by both the physical structure of an individual's vocal tract and the behavioral characteristics of the individual, for identification, verification or other related tasks.

speech recognition

A technology that enables a machine to recognize spoken words. Speech recognition is *not* a biometric technology that identifies an individual based on physical and behavioral characteristics (see speaker recognition).

spoof / spoofing

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

stand alone resources

SIV and ASR are separate engines whereby each engine returns its own results. The term is an antonym of co-located resources.

supervised adaptation

See adaptation.

template

A term used by the biometrics industry. Sometimes referred to as reference model, voice model or voiceprint.

text dependent

An SIV technology (usually verification technology) that 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). During verification, the user is prompted to speak the pass-phrase they enrolled with. The verification session depends on the designated pass-phrase (spoken by a particular user) in order to function properly in *text dependent* mode.

text independent

SIV technology that can operate 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. During verification, the user is prompted to speak any text regardless of what they spoke during enrollment. The verification session does not depend on the particular text but does require an adequate amount of input speech by the user in order to function properly in *text independent* mode.

text prompted

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. One simple example is when a user has been enrolled via voice data collected through a series of prompts to speak the answers to a number of identity related question. During a verification session, one of the identity questions is selected in random order (What is your pet's name?). The user is then prompted to speak the answer to the selected question.

threshold

The threshold is 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.

true speaker

The true speaker is a person whose identity is the same as their real identity. Antonym of imposter.

turn

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

unsupervised adaptation

See adaptation.

utterance

An utterance is a spoken input speech sample. It 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

Speaker verification is a biometric modality that uses an individual’s speech for verification of a claimed identity. Verification is a task where the biometric system uses an *identity claim* for reference *matching* against a submitted biometric sample (processed voice sample) and if matched returns a positive verification. A processed voice sample is collected and compared to one reference model or a small group of reference models in a database. See *identity claim* for examples. Verification is also referred to as one-to-one matching.

voice model

It is a system’s representation of an individual’s voice and is constructed from data extracted from one or more voice samples provided by that individual. A voice model is created during verification, identification or enrollment. Voice models created as an outcome of enrollment are synonymous with the preferred term *reference model*. Also referred to as voiceprint or template.

voice sample

Spoken input that may consist of one or more utterances.

voiceprint

Voiceprint can also be referred to as a reference model, voice model or template.