Configuration and Management of Speaker Verification Systems

Chuck Johnson
Architect
iBiometrics, Inc.

Presentation for the WC3 Workshop on Speaker Biometrics and VoiceXML 3.0

The performance (accuracy) of Speaker Verification (SV) systems is a function of how well the engines are configured and managed, and how well the verification results are interpreted.

Speaker Verification resources (engines) are initially configured (setup) using some set of parameters to yield optimal results for a broad set of users in common application scenarios.

For peak performance of a Speaker Verification solution, the VoiceXML client (voice application) needs to be able to query and set the necessary initialization and configuration (setup) parameters, and control the operation of Speaker Verification resources.

This paper will present, discuss and recommend configuration/initialization parameters and operational controls that should be available to the VoiceXML client.

Configuration and Management of SV Resources

Modes of operation

Many Speaker Verification engines support multiple modes or types of verification – text dependant, text independent, and/or text prompted.

The VoiceXML client should be able to query the engine to determine the available mode(s) of operation. The client should be able to select/set the available mode(s) of SV operation.

Setting operating points, verification thresholds or security levels

Setting the operating point of an SV engine will establish the False Match (False Accept) rate and False Non-match(False Reject) rate. The operating point of an SV engine (the match and non-match rates) can be set from results of a 'tuning exercise' that uses some number of utterances from a representative sample of users. The exercise generates a ROC curve along which an operating point can be selected and set.

Some SV engines support the setting of 'verification thresholds' or 'security levels'. The setting of these levels establishes the operating point - the minimum False Accept rate and, optionally the False Reject rate, for the SV engine. Depending on the SV engine, the False Reject rate may or may not be established (not specified or guaranteed). When using these settings some SV vendors advise that "Specific [verification] performance is highly dependent on the verification dialog and may thus fall outside of the specified range".

The VoiceXML client should be able query the operating point (verification threshold or security level). The client may be able to set the operating point.

Setting the operating point for different user groups or populations

Many SV applications can have distinct user groups or populations. Each user groups could have different security policies. Those policies could recommend/require different operating points, verification thresholds or security level settings.

The VoiceXML client should be able to set/establish the SV engine operating point for a specific user group or population.

Selection and use of specific background models

Different user groups (populations) may have group specific background (imposter) models or specific cohort models.

The VoiceXML client should be able to load and utilize the appropriate user (group) background model or cohort model.

Configuration for multi-utterance verification (within a single verification session)

SV engines may support multi-utterance verification sessions – capture and analysis of two or more utterances to render a single, final verification decision (e.g. Nuance variable-length verification).

The VoiceXML client should be able to query and set multi-utterance verification parameters (i.e. minimum and maximum number of utterances). The client may be able to query intermediate (cumulative) verification results.

Voice model adaptation settings (un-supervised)

Un-supervised adaptation (updating the voice model) is automatically performed by the SV engine on a predetermined basis (i.e. when the verification scores exceed a preset threshold. Note: This threshold is usually higher than the verification pass/fail threshold).

The VoiceXML client should be able to query the engine to determine if un-supervised adaptation has been enabled/disabled, and obtain threshold setting information. The client should be able to enable/disable adaptation. The client may be able set the adaptation threshold(s).

Voice model adaptation (supervised)

Supervised adaptation (updating the voice model) is performed by VoiceXML client application based on application specific criteria.

The VoiceXML client should be able to request adaption of the voice model.

World model adaptation

The world model (depending on literatures and usage may also be known as background or imposter model) is built from a large number of representative speakers.

The VoiceXML client may be able to update the world model.

Setting required phrase

Text dependant verification requires the voice input of one or more specific pass-phrases.

The VoiceXML client should have the capability to set the required (pass) phrase.

Enrollment

Enrollment is the process of collecting voice samples from a person and the subsequent generation and storage of voice reference modes (voice model, voiceprint) for that person.

The VoiceXML client should be able to determine if the SV engine is ready for training. The client should be able to set the required (pass phrase) – for text dependant verification. The client should be able to query the engine to determine the enrollment status of a user’s voice model. The client may be able to query the engine to determine the minimum amount of training audio required (and maximum amount of audio allowed), or minimum and maximum number of training utterances.

Utterance length (min and max) to support different types of speaker verification

SV engines require some [absolute] minimum amount of audio to perform text independent, text dependant, or text prompted verification. The engine may recommend an amount of audio for optimal (best) results. The engine may set limits on the maximum amount of audio.

The VoiceXML client may able to query the engine to determine the minimum, optimal and/or maximum amount of audio for each type of verification that is supported.

Rollback

Rollback is a mechanism to ‘undo’ or roll back the processing performed on the last turn (and only the last turn) in an enrollment or verification session.

The VoiceXML client should be able to request the rollback of the processing that was performed during the last turn.

Buffering

The audio (used for enrollment and verification) may be streamed to the SV engine, or may be delivered as a file or a shared buffer.

The VoiceXML client should, at a minimum, be able to access the audio that was captured during the current/last turn of a verification session. The client may be able to request the SV engine perform additional verification or enrollment (training) tasks/activities.