CalSAWS

California Statewide Automated Welfare System

Design Document

CA-239812

Voice Biometrics Sample Collection

CalSAWS	DOCUMENT APPROVAL HISTORY	
	Prepared By	Jared Kuester
	Reviewed By	Darcy Alexander, Logan Pratt, Michael T. Wright

DATE	DOCUMENT VERSION	REVISION DESCRIPTION	AUTHOR
1/21/2022	1.0	Initial Draft	Jared Kuester

Table of Contents

1 OVERVIEW

To support additional languages in the Voice Biometrics System (CA-226843), 200 voice recordings samples (100 male and 100 female) need to be collected for each new language. These recordings will be used to create a voice model for each language.

1.1 Current Design

The Voice Biometrics solution currently only supports English and Spanish language.

1.2 Requests

Create a new IVR to collect sample recordings and track how many recordings we have gathered for each language and gender of the recordings. Modify the Login Results Contact Flow in each IVR to query the preferred spoken language of the customer and request if the customer would like to ==record three voice== samples.

1.3 Overview of Recommendations

1. Create a new inbound IVR to collect voice samples
   a. Request the customer to select their language
      i. Farsi
      ii. Vietnamese
      iii. Mandarin
      iv. Tagalog
      v. Russian
      vi. Korean
      vii. Cambodian
      viii. Hmong
      ix. Arabic
      x. Lao
      xi. Cantonese
      xii. Armenian
      xiii. Portuguese
   b. Request the customer to select their gender identity
      i. Male
      ii. Female
      iii. Non-Binary
   c. Query a new DynamoDB to see if we still need a recording for the selected language and gender
   d. Prompt the customer to say the phrase "My voice is my password, please verify me" in their selected language three times
2. Modify the Login Results Contact Flow in existing Amazon Connect inbound IVRs.
   a. If customer successfully authenticates, query CalSAWS database for their preferred spoken language and gender Identity

b. Query the DynamoDB to see if a recording for that language and gender identity is still needed
c. If more are needed, request the customer in their preferred spoken language to make three recordings
d. If the customer opts in, prompt them to say the phrase "My voice is my password, please verify me" three times.

1.4 Assumptions

1. Inbound IVR will only prompt customer to record their voice if additional recordings are still needed.
2. The pool of recordings is statewide and across both recording methods.

2 RECOMMENDATIONS

2.1 New Inbound IVR

2.1.1 Overview

Create an IVR used to collect voice samples that will be used in creating the new Voice Models for CA-226843. Without these samples, a voice model can't be created. One hundred male, and one hundred female voices in each new language is required to create the new voice model.

2.1.2 Description of Changes

1. Create a new IVR Call Flow used to record your voice.
   a. When the customer first calls in, they are prompted to select one of the following languages.
      i. Farsi
      ii. Vietnamese
      iii. Mandarin
      iv. Tagalog
      v. Russian
      vi. Korean
      vii. Cambodian
      viii. Hmong
      ix. Arabic
      x. Lao
      xi. Cantonese
      xii. Armenian
      xiii. Portuguese
   b. After selecting a language, they are prompted to select their gender identity.
   c. The IVR will then check if a voice recording is needed in the selected language and gender identity.
   d. Store the voice sample in an S3 bucket and update the database that a successful recording has been captured.
   e. For more information, please see the attached call flow (Voice Sample Call Flow.pdf).

2.2 Update Inbound IVR Login Results

2.2.1 Overview

To collect voice samples from the public, update the Login results contact flow to query the CalSAWS database for the customer's preferred spoken language, and their gender identity. If we still need a voice sample in that language and gender identity, ask the customer if they would like to participate in an upcoming change to the IVR.

2.2.2 Description of Changes

1. Modify the Login Results lambda to query the CalSAWS database upon successful login
   a. Collect the following information
      i. Spoken Language
      ii. Gender
   b. Query the same database used in the new IVR to see a recording in that language and gender identity is still needed.
   c. If yes, prompt the customer to participate by recording the phrase, "My voice is my password, please verify me" in their preferred language three times.
   d. After successful or failed recording, return the customer back to the normal place in the IVR.
   e. For more information, please see the attached call flow (Voice Sample Call Flow.pdf).

3 SUPPORTING DOCUMENTS

Number	Functional Area	Description	Attachment
1	IVR	Call flow for both the New IVR and the update to the existing Inbound IVRs.	Voice Sample Call Flow.pdf

4 REQUIREMENTS

4.1 Project Requirements

REQ #	REQUIREMENT TEXT	How Requirement Met
DDID 2701	The CONTRACTOR shall configure the CalSAWS Inbound IVR to allow a customer to enroll and capture their voice print for voice authentication.	Collect Voice Samples to create the background voice models. These voice models will be used to complete SCR CA-226843.

5 APPENDIX