Speech Database/Tool System and Preliminary Accent Recognition Study

Arthur Phidd, Padmashree Thimmappa, Richard Sauther and Shankar Vijayakumar
Under the guidance of Professor Charles Tappert

Abstract

The voice/tool database System is a web-based system connected to a relational backend database that stores voice files (.wav format) along with user demographic information. The database is used to test the ability of an examiner, usually a briefly trained non-linguist, to identify the nationality of the speaker of the recorded voice. Here we provide background on voice recognition biometrics, describe the design and functionality of the pronunciation information system, and present some preliminary results on the nationality classification of voice samples by examiners. We also provide the ability to play the voice files in a speech processing tool for further research in the field.

1. Introduction

Biometrics is a field that recognizes the differences in people's physical makeup and behavioral characteristics and identifies people based on those differences [1] [2]. The voice biometric technology analyzes the tone, pitch, cadence, and frequency of a person's voice. Using the distinct speech patterns that are common in individuals from different regions of the world, we are developing a tool that can verify the nationality specific pronunciation claims made by language experts. Target words and phrases are incorporated into a specially designed text that is read and recorded by individuals from various national backgrounds. Our system has the ability to measure the accuracy and skills of the examiners who analyze the recordings, whether they are trained linguists or briefly-trained non linguists, to estimate the speaker's correct nationality from database of voice files with known identities.

2. Design and Implementation of the Voice Recognition System

The user interacts with the system through a web interface that is connected to a backend database. A subject's recorded speech is saved as a .wav file in the database along with the subject's demographic information. Once samples have been uploaded into the database, the examiner (user) is evaluated on his/her ability to recognize regional pronunciation distinctions and to classify a sample based on its nationality. The primary components of the required product include:

A web-based system connected to a relational backend database to capture the user's demographic information and a voice sample in a .wav format.

A Pronunciation Affinity Matrix (PAM) designed by the customer has been incorporated into the system's web interface to assist the user with classifying voice files. The user will employ the PAM to score the target words and the grading algorithm will evaluate the user's ability to select the correct nationality. Each word pronunciation will be given an affinity score between 0 and 5 with 5 being the highest level of affinity confidence.

The following user stories were sequentially implemented.

Story#1: Enroll voice donors

We created a screen that the users will use to input their reference information using a list of give input fields and a second screen that allows the user to upload their .wav format. The screen should browse the user's hard drive allowing the user to select a file. Files are saved to a directory on the server and

given a file name corresponding to the user's computer-generated unique ID in the database. A table holding reference information for each speaker is created. The DFD for Story 1 is shown in Fig 1.
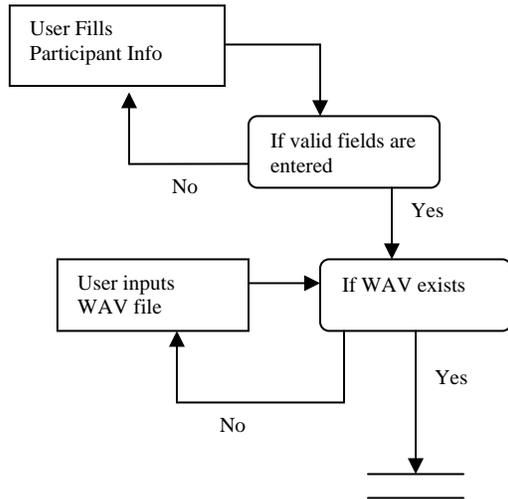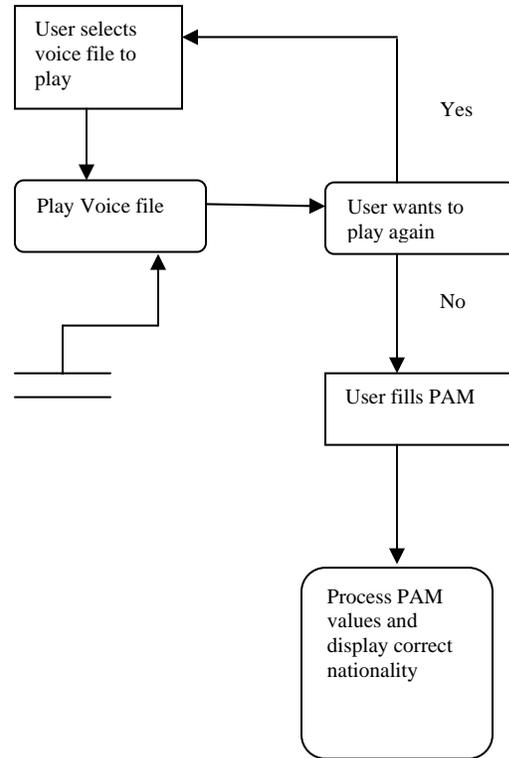

Fig1: DFD for Story#1.

Fig 2: DFD for Story#2.

Story#2: Determine Speaker's Nationality or Regionality

A screen was created which the users will use to select a voice file, play it back and determine the Nationality of origin using the Pronunciation Affinity Matrix (PAM).The DFD for Story 2 is shown in Fig 2.

Use Cases and User Interface Prototyping

1. Voice Lab
Primary actors
- Voice donor – individual who uploads a voice sample and personal demographic information to the database.
- Voice evaluator – an individual who plays a recorded voice sample and will use the PAM and its suggested pronunciation patterns to determine the nationality of the subject.

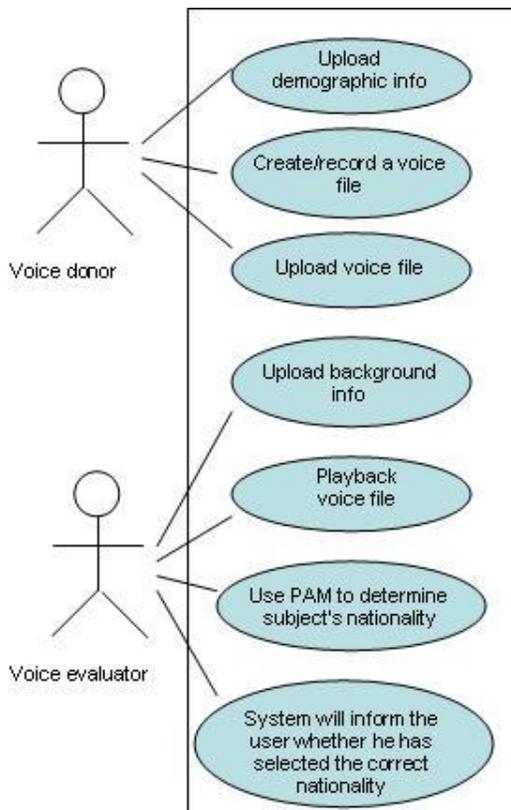Fig 3 shows the use case diagram for the voice lab.

Fig 3: Use case Diagram for Voice Lab

1.1 Use Case: Upload Demographic information
Primary actor: Voice Donor

Goal in Context: To enter demographic information into the web interface that provides a description of the subject. Required fields include last and first name, birth-county, native language, etc . . . When user hits submit, all input fields are uploaded to the database and the user advances to the voice file upload screen.

Scenario: User observes control panel, enter valid input data, presses submit to transmit data to database.

Exceptions: All input fields must have valid entries.

1.2 Use Case: Create/Record a voice file
Primary actor: Voice Donor

Goal in Context: Using the voice recording software present on the user's pc, record the selected text as a .wav file.

Scenario: User sitting in front of a PC with a microphone records his/her voice and saves the file in a .wav format.

Exceptions: Desired system users may not have access to microphone or audio recording software.

1.3 Use Case: Upload a voice file
Primary actor: Voice Donor

Goal in Context: Upload the saved voice file from a drive on the user's PC to a directory on the server and save the file location to the project database. File will be stored and associated with the previously uploaded demographic information.

Scenario: After the user enters all valid demographic data and submits, he/she will advance to the file upload screen. User will click the browse button to search and choose the .wav files from their PC. Once a file is selected, user hits submit and the file is uploaded to the server and the file location saved in the project database.

1.4 Use Case: Upload voice evaluator's demographic information
Primary actor: Voice Evaluator

Goal in Context: To enter demographic information into the web interface that provides a description of the person who will evaluate the voice file. Required fields include last and first name, birth-county, native language, etc . . . When user hits submit, all input fields are uploaded to the database and the user advances to the Pronunciation Affinity Matrix (PAM).

Scenario: User observes control panel, enter valid input data, presses submit to transmit data to database.

Exceptions: All input fields must have valid entries.

In summary, the user picks a voice file from a list box.

User clicks a play button to hear recorded speech. User tests the recording for each word in the American English column and enters an affinity score in the score column. This process is repeated for each language in the Matrix and when all are scored, the matrix will calculate the overall rating and

tell the user the correct nationality or Region.

Story#3: Provide a tool for user to view voice measurement data

Several speech processing tools are available for experimental speech research in such areas as speech recognition, speaker authentication, and voice biometric studies. In order of importance to the customer, these tools include:

•A speech spectrogram tool to perform a spectral analysis of a speech signal. This is a standard speech visualization tool that typically gives a grey-scale plot of frequency bands as a function of time.
•A procedure to facilitate the recognition of a person's nationality from his/her speech patterns. For example, an interactive, speech-data quiz might be created to measure the nationality recognition accuracy of trained linguists or briefly-trained non-linguist examiners.
•A tool to segment a speech portion of the signal from the background noise by having a threshold of the signal's energy function.
•A tool that uses the elastic matching (dynamic time warping) algorithm to align a speech signal with a pre-segmented one for alignment purposes.
A search was made to find such tools on the Internet. The tools should be incorporated in a directory, possibly to be used interactively through the Web interface. The spectrographic tool should provide access to the actual numerical data (e.g., the energy in a particular frequency band in a particular time interval) that can be processed later in an application. This tool should also include functionality to separate speech from background noise. Our research yielded several results for speech processing tools with the above mentioned capability. We have found that one of them is serving our purpose and is less complicated than the other software's.
This is the Speech Filing System (SFS) [6].

Overview of SFS: SFS 4/Windows is a free computing environment for PCs for conducting research into the nature of speech. It comprises software tools, file and data formats, subroutine libraries, graphics, special programming languages and tutorial documentation. [7] It performs standard operations such as acquisition, replay, display and labeling, spectrographic and formant analysis and fundamental frequency estimation. It comes with a large body of ready made tools for signal processing, synthesis and recognition, as well as support for your own software development.

Implementation
The application is implemented as seen below:
 http://utopia.csis.pace.edu/cs615/2004-2005/team3/Story4.htm takes you to the initial screen of the voice lab. As shown in Fig 4, the initial screen has links to submit the voice files, Analyze the Voice files by submitting the voice file and then downloading the Speech Filing System to install and execute the software to further obtain numeric data of the voice file.
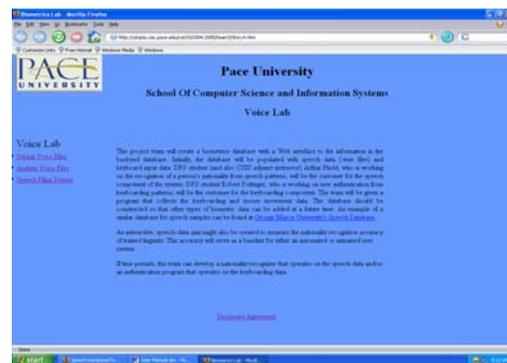

Fig 4: Voice lab

The Submit voice file link takes you to the Participant Form as shown in Fig 5. The user can enter values into the various fields, select a voice file and submit the form. Once he submits the form all the data will be recorded in a back end database. The uploaded voice file can be later used for research purposes.

Fig 5: Participant Form

Once you submit the information. Your data entry is confirmed and you can choose to either analyze your voice file or use the Speech Filing System for analysis of voice files.

The screen shot for the successfully entered values in the database is as in Fig 6.

The participant can view his User ID and the name of his file which he has submitted

Fig 6: Data input information.

The Analyze Voice Files link takes the user to the Voice Analysis screen as shown in Fig 4 where he can play a Voice file and analyze it. He can give a score to each of the key words in the voice file which ranges from 0 to 5. This range is implemented as a drop down list. The score for each word can be selected and can be submitted. Fig 6 shows this:
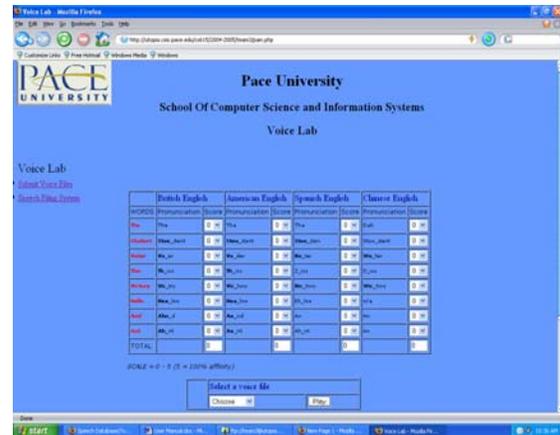

Fig 6: Analyze voice files

Once the above file is submitted the possible nationality of the participant is determined. The next screen as in Fig 7 shows the nationality of a person and his information retrieved from the database.
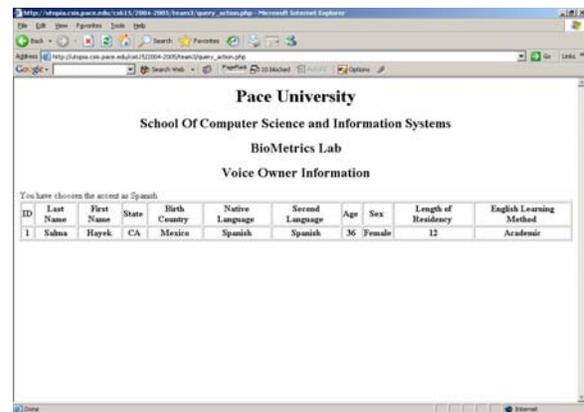

Fig 7: Voice owner Information

Speech Filing System link gives the user an ability to run his voice files through a voice analysis software called the Speech Filing System. As implemented at this stage, this page allows the user to install the software first and then feed his voice files to this system. Also if he has any problems accessing the Installation link, he can visit the website of SFS through the link mentioned on the below to download the software and install it by himself. The screen is as shown in Fig 8

Fig 8: Analysis with SFS

After SFS is installed, voice files can be analyzed using the wide spectrogram as shown in Fig 9.
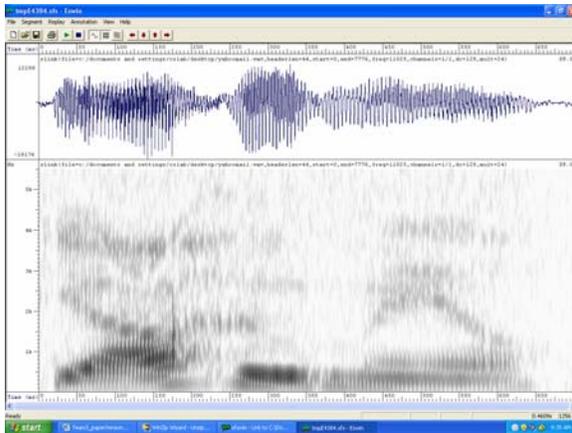

Fig 9: Wide spectrogram of a .wav file

Testing

Testing For Webpage Validations and database entries:

Validations have been completed for two PHP pages namely Participation.php and voicereg.php.

Participation.php: This form takes in the information of the participant who would wish to register his voice data.

Voicereg.php: The form takes in a voice file of the participant.

The Participation form is tested for the following validations:

1. Every field is mandatory and cannot take null values.

2. Fields cannot have special characters.

3. Fields like First name, Last name, Country, Native language, Current language, Current state cannot have numerical characters in the entry.

4. The length of each field is checked for their respective length constraints:

The length of the First name cannot be less than 1 character.

The length of Last name cannot be less than 2 characters.

The length of age cannot be 0 or negative.

5. Year of residence cannot be greater than the age.

6. The form is also checked to see if the participant has entered a filename.

7. When the user submits data, the database is checked if the voice file name matches the User ID of that Participant.

8. The Personal Affinity matrix has been checked to indicate the right sum for the values chosen by the user for the key words.

9. Testing has been done to determine the nationality of a participant.

The application is also tested for playing appropriate voice files and displaying the voice files in the database through a drop down list. The Accent chosen is also tested to take in values and give the right total of the selected values.

4.   References:

[1] John D Woodward Jr, Peter T Higgins, Biometrics, Identity Assurance in information Age, April 2003.

[2]Julian Ashbourn, Implementing Biometric Security, March 2004.

[3] Net Nanny Software International, Inc.Technical Report BioPassword Keystroke Dynamics. Net Nanny Software International, Inc. 2000-2001
http://www.biopassword.com/home/technology/ BP%204.5%20Technical%20Paper.pdf

[4] Beginning PHP 5 from Novice to Professionals by W.Jason Gilmore.

[5] Professional PHP Programming by Jesus Castagnetto.

[6] SFS Online Manual
http://www.phon.ucl.ac.uk/resource/sfs/help/index.html

[7] SFS Software available at:
ftp://ftp.phon.ucl.ac.uk/pub/sfs