## **Singing Data Labeling Tool**

Carlos Cepeda - ccepeda2018@my.fit.edu

Avinash Persaud - apersaud2018@my.fit.edu

Nandith Narayan - nnarayan2018@my.fit.edu

Faculty advisor: Dr. Shoaff

## Client:

The client is the general vocal synthesis community and voice technology researchers, however we will communicate with a small group of users within the community to assure our goals are being met.

Caleb Matthew Long from Appalachian State University will act as a representative of the target group.

Date(s) of Meeting(s) with the Client for developing this Plan: 9/2/21

## Goal and motivation:

Our goal is to create a user-friendly tool to assist the client with labeling singing data for machine learning purposes. The current tools for labeling singing data that exist are sub-par; multiple tools are often required to label singing data.

## Approach (key features of the system):

- Creation of Mono Labels. Mono Labels are a special label format that contains information about the data; this includes starttime, endtime, and a phoneme identifier. This allows the user to tag and label singing data as well as align phonemes, all in the same tool.
- Creation of extra feature layers. A feature layer contains a specific representation of the data, for example: notes, syllables, and formants are all examples of feature layers.
- Exporting to multiple output templates including the HTS Full Label format. This format has two variations: Score aligned, and Phoneme aligned.
- Automatic phoneme alignment and detection. This serves to reduce the amount of work the user would need to do. It is quicker and more accurate than manually aligning every phoneme.
- Storing an intermediate representation in a project file. This allows the user to quickly pick up where they left off. It also allows for the export of the data into multiple formats.
- User-Interface that provides the client easy access to the different editing tools, and a convenient way to display the singing data for easy understanding.

Novel features/functionalities:

• The ability to copy and paste phonemes. The current tools don't allow for easily copying and pasting phonemes.

• Custom output templates.

Technical Challenges:

- One of our target output formats (HTS) has a lot of specifications and is challenging to correctly generate. Creating a customizable output for this format presents a challenge.
- Automated phoneme detection and alignment.
- The User Interface will contain a lot of elements and we are not very experienced with UI design and programming.

Milestone 1 (Oct 4): itemized tasks:

- For automatic phoneme alignment, we would like to compare 'Shiro', 'CMU Sphinx', ...
- Provide small demo(s) to evaluate 'Shiro', 'CMU Sphinx', ...
- For creating the User Interface, we plan to compare the following tools: WXwidgets, Qt, GTK+
- Provide a sample GUI made in each of the tools (WXwidgets, QT, GTK+)
- For generating and parsing the output format, we plan to compare Antlr, ...
- Resolve technical challenges: Creating a general User Interface.
- Compare and select collaboration tools for software development, documents/presentations, communication, task calendar
- Create Requirement Document
- Create Design Document
- Create Test Plan

Milestone 2 (Nov 1): itemized tasks:

- Implement, test, and demo general GUI layout.
- Implement, test, and demo automatic phoneme alignment.
- Implement, test, and demo the intermediate representation and structure of the data.
- Implement, test, and demo the saving and loading of project files.

Milestone 3 (Nov 29): itemized tasks:

- Integrate the User Interface with the intermediate data structure
- Implement, test, and demo the spectrogram and graph elements of the GUI.
- Implement, test, and demo inputting features into the tool.

Task matrix for Milestone 1

Task Car	arlos	Nandith	Avinash
----------	-------	---------	---------

Compare and select Technical Tools	GUI	Parsers Phoneme Alignment	Phoneme Alignment
"hello world" demos	GUI	Parser	Phoneme Alignment
Resolve Technical Challenges	GUI layout planning	Algorithm and tool for parsing and generating output format.	Selecting tool and making datasets as needed
Compare and select Collaboration Tools	Version Control	Documents and presentations	communication, task calendar
Requirement Document	write 25%	write 25%	write 50%
Design Document	write 34%	write 33%	write 33%
Test Plan	write 25%	write 50%	write 25%

Approval from Faculty Advisor

- "I have discussed with the team and approve this project plan. I will evaluate the progress and assign a grade for each of the three milestones."
- Signature: \_\_\_\_\_ Date: \_\_\_\_\_