Singing Data Labeling Tool Plan

Team: Nandith Narayan, Avinash Persaud, Carlos Cepeda
Advisor: Dr. William Shoaff

The problem

Vocal synthesis is the process of creating artificial human speech. Just like any other machine learning process, vocal synthesis requires a lot of data. Not only does this require data, but it also requires that data to be labeled. Manually labeling this data takes up a large amount of time and effort.

Goal and Motivation

We aim to develop an integrated and efficient user interface and set of tools to assist with labeling singing data for vocal synthesis purposes. The current tools available have a variety of issues.

Current tools have the following issues:

- Not one tool does all tasks
- Interfaces are unintuitive
- Labor intensive

Types of Labels

There are numerous ways to label singing data, a few of the most common methods are listed below:

- Phonemes (individual consonant or vowel sounds in a language that come together to form words)
- Syllables
- Notes

Our Approach

1. Provide convenient UI features for labeling features (Phonemes, Syllables, Notes, etc.)
    - Shortcuts for common actions - creating, deleting, copying, and pasting of Phonemes.
    - Fast and fluid navigation - placing buttons and dropdown menus for key tasks in easy to reach parts of the user interface.
    - Informative display - provide the user with information about the data such as the Fourier transform of the data.

Our Approach

2. Automated tools to allow extrapolation between features
    - Automatically extrapolating syllables and phonemes from the music score and lyrics.
    - Automatically extrapolating syllables and notes from phonemes.

Our Approach

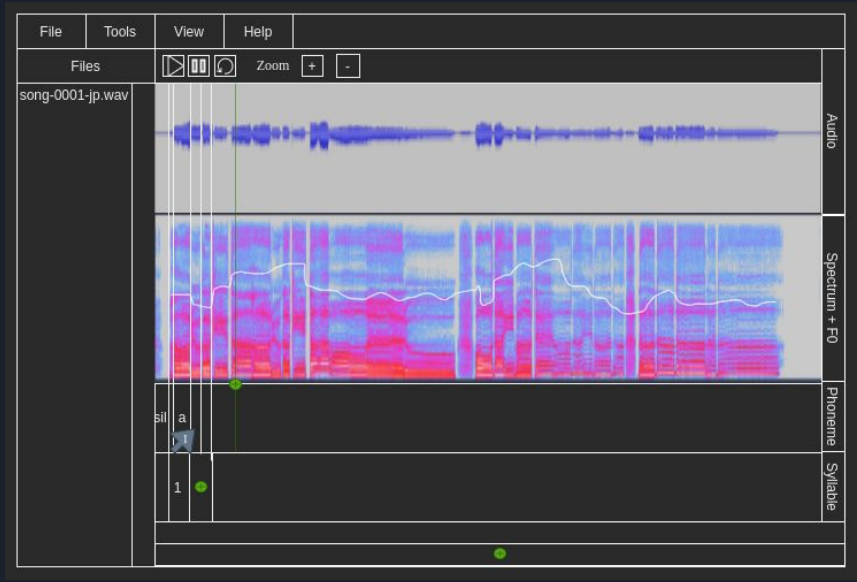
3. Tools for automatically aligning/time marking location of phonemes
    - Automatically detecting the phonemes present in the audio.
    - Automatically align the start and end of the phonemes with the audio data.

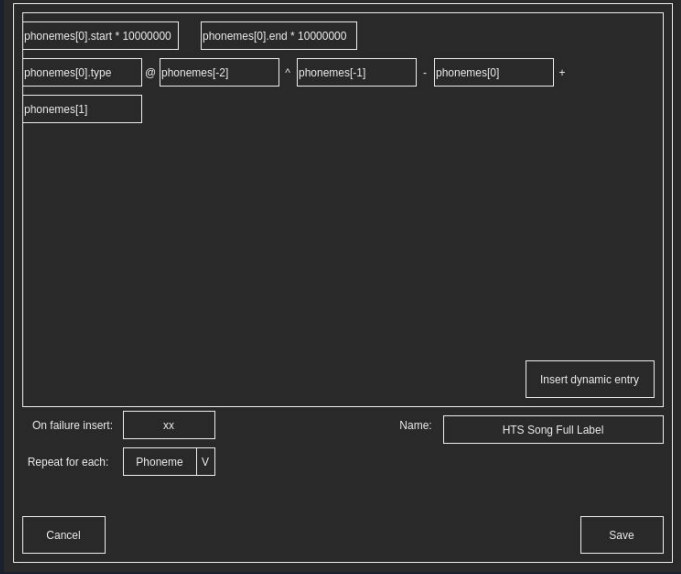
4. Ability to configure custom outputs
    - Allow the user to select an output format to export the labeled data to.
    - We are targeting the HTS singing label format.

Technical Challenges

- The HTS Singing Label format is a complicated output format to cover
  - This output format is very dense which makes it difficult for a human to read and parse.
  - Verification of the correctness of the output.
- Finding and integrating, possibly developing, a tool for phoneme alignment which operates cross platform
  - Most automated tools are command line based and don't have unified integration with other steps of the process
- Creating an intermediate data structure to store the project data
  - This data structure must be able to be used to generate the output in the various formats.
  - This data structure must preserve the state of the labeling progress so that the user may close the program and resume labeling later.
  - This data structure must save and load quickly.

Sample UI layout





Milestone 1

October 4th

- Compare, select, and create “hello world” examples of tools for phoneme alignment, parsing for dynamic output, and graphics library
- Compare and select collaboration tools for software development, documents/presentations, communication, task calendar
- Resolve technical challenges that are presented
- Create Requirement Document
- Create Design Document
- Create Test Plan

Milestone 2

- Implement, test, and demo phoneme alignment, user interface, intermediate data structures, and saving/restoring data

November 1st

Milestone 3

- Integrate, test, and demonstrate the user interface's ability to use the intermediate data structures
- Implement spectrogram and graph elements
- Allow for inputting, saving, and restoring feature layers and labels

November 29th

Task Matrix For Milestone 1

Task	Carlos	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%