Crowdsourcing the Audio Transcription of Songs: Designing a User Interface to Facilitate Large Scale Audio Data Correction

I. Avendano, J. Allison, & D. Shanahan
Center for Computation & Technology

Introduction

In the field of music information retrieval (MIR), it can be difficult to provide access to transcriptions for large collections of songs. To address this challenge, researchers look towards algorithms that output automated transcriptions of audio files. However, current MIR algorithms often output inaccurate automated transcriptions with errors such as miscalculated pitch, tempo, and note duration.

This project provides a solution to this problem through a gamified user interface. Users can modify and improve the transcription of an audio file through an interactive Piano Roll. The idea behind this interface is to combine computing abilities and crowdsourced human interest to transpose songs as accurately and efficiently as possible.

Tools

1. Libraries used include:
   - D3 (Data Driven Documents) used to visually display and make modifications to the audio data
   - Tone JS library to control real-time web audio
2. MP3 to midi file conversion done using:
   - Sonic Visualizer Silvet Note Transcription Plug-In
   - Melodyne
3. Midi to JSON file conversion done with ToneJS Midi Convert Demo. Ex. Of Note Object:

```
{
    "name": "C1",
    "midi": 24,
    "time": 2.6546681666666663,
    "velocity": 0.5826771653543307,
    "duration": 3.639089145833333
},
```

4. Demo dataset includes 50 audio versions of the National Anthem sung at the past 50 Super bowls
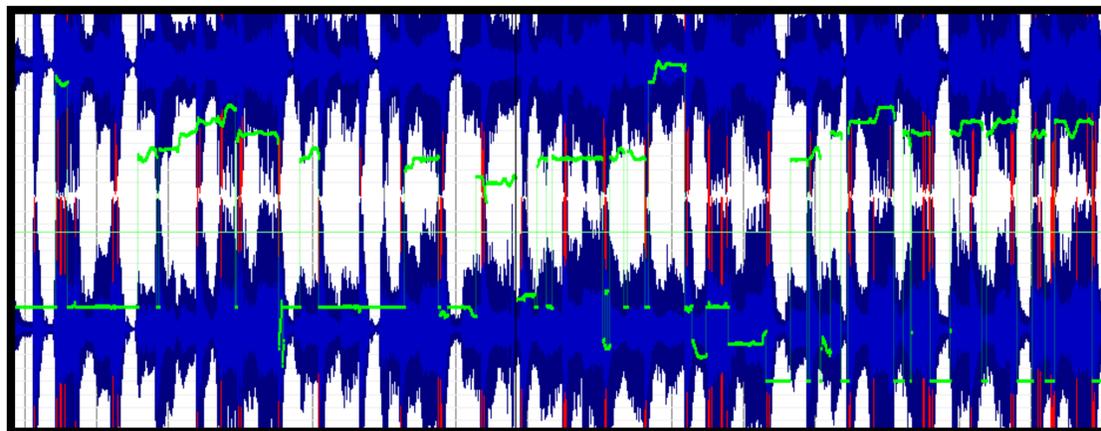
Piano Roll

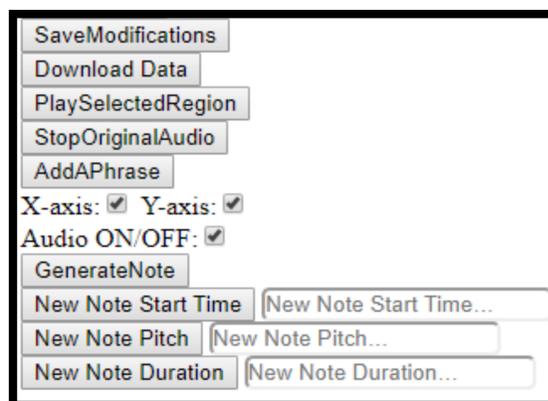Figure 1: Sonic Visualizer Automated Transcription

Figure 2: Piano Roll User Interface

Modification features on the piano roll:
- Drag notes to change "time" key value1
- Resize notes and change "duration"
- Add and Delete notes
- Drag note on y-axis to change pitch
- Create array of regions that designate national anthem lyric/phrase to a section
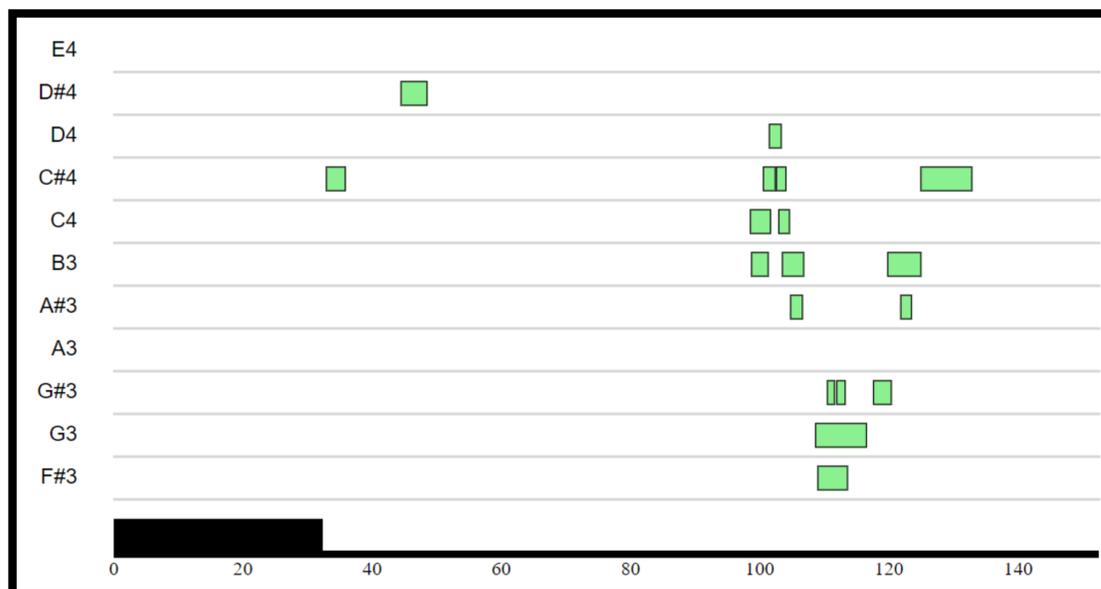- D3 brush selects section of notes and section of original audio file for playback

Figure 3: Piano Roll with partial transcription selected. Notes enlarged to scale

Future Directions

With enough recurrent user-modifications of the same audio file, the CATS program will analyze the modified transcriptions, eventually determining the correctly transcribed version of the audio file. The data of the modified transcribed audio will then be stored in an accessible library of accurate digital music transcriptions.

The goal is to eventually analyze patterns in human-made modifications of the same song and train a MIR algorithm to transcribe songs like a human user does, minimizing pitch, tempo, and note length miscalculations in automated transcriptions outputted by MIR algorithms.

References

"Sonic Visualizer Silvet Note Plug-In." *Vamp Plugins*. Queen Mary, University of London

"Sonic Visualiser." *Sonic Visualiser*. Centre for Digital Music, Queen Mary, University of London

Bostock, Mike. "Data-Driven Documents." *D3.js*. N.p., n.d. Web.

"Tone.js." *Tone.js*. N.p.

"D3 Event Action At A Distance." *DashingD3js.com*. N.p

Acknowledgements