Reproducibility Companion Paper: MeTILDA - Platform for Melodic Transcription in Language Documentation and Application

Mitchell Lee
Computing & Software Systems
University of Washington Bothell
Bothell, WA, USA
mitchl2@uw.edu

Chris Lee
Computing & Software Systems
University of Washington Bothell
Bothell, WA, USA
leec351@uw.edu

Sanjay V Penmetsa
Computing & Software Systems
University of Washington Bothell
Bothell, WA, USA
penmetsa@uw.edu

Min Chen*
Computing & Software Systems
University of Washington Bothell
Bothell, WA, USA
minchen2@uw.edu

Mizuki Miyashita
Department of Anthropology
University of Montana
Missoula, MT, USA
mizuki.miyashita@umontana.edu

Naatosi Fish
Blackfeet Community College
Browning, MT, USA
naatosifish@gmail.com

Bo Wu†
MIT-IBM Watson AI Lab
Cambridge, MA, USA
bo.wu@ibm.com

Omar Khan‡
Reykjavik University
Reykjavik, Iceland
omark@ru.is

ABSTRACT

This companion paper supports the replication of the development and evaluation of "MeTILDA - Platform for Melodic Transcription in Language Documentation and Application" that we presented in the ICMR 2021. MeTILDA aims to help document and analyze pitch patterns of endangered languages including Blackfoot, whose prosodic system is characterized by pitch movements. It develops a new form of audio analysis (termed MeT scale which is a perceptual scale) and automates the process of creating visual aids (Pitch Art) to provide more effective visuals of perceived changes in pitch movement. In this paper, we explain the file structure of the source code and publish the details of our data as well as system operations. Moreover, we provide a link to the demo video for facilitating the use of our platform.

CCS CONCEPTS

• **Applied computing** → **Computer-assisted instruction**; • **Information systems** → **Data analytics**.

KEYWORDS

Endangered language; MeTILDA; Perceptual scale; Pitch Art

*Corresponding Author.
†Reproducibility Reviewer.
‡Reproducibility Reviewer.

ICMR '23, June 12–15, 2023, Thessaloniki, Greece

ACM ISBN 979-8-4007-0178-8/23/06.
https://doi.org/10.1145/3591106.3592277

ACM Reference Format:
Mitchell Lee, Chris Lee, Sanjay V Penmetsa, Min Chen, Mizuki Miyashita, Naatosi Fish, Bo Wu, and Omar Khan. 2023. Reproducibility Companion Paper: MeTILDA - Platform for Melodic Transcription in Language Documentation and Application. In *International Conference on Multimedia Retrieval (ICMR '23), June 12–15, 2023, Thessaloniki, Greece.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3591106.3592277

1 ARTIFACTS DESCRIPTION

1.1 Introduction

In our original paper [2], we presented a system called MeTILDA (Melodic Transcription in Language Documentation and Application) for endangered language analysis and preservation. It provides new forms of audio analysis for pitch accent languages such as Blackfoot [1], and automates the process of creating visual aids called Pitch Art to teach the nuance in pitch changes. Specifically, a perceptual scale is developed to reflect how pitches are auditorily perceived by human using a variation of the western music scale. This scale provides a common reference for comparing pitch across recordings, regardless of the speakers' natural pitch range. In addition, MeTILDA consolidates and automates the workflow of generating Pitch Art for Blackfoot teachers and learners to understand how their pronunciation compares to that of native speakers. It also helps linguistics researchers in their efforts to document and transcribe audio recordings of endangered languages. In this companion paper, we present a replication artifact that provides a complete re-implementation and evaluation of the MeTILDA system discussed in the original paper.

1.2 Source Code Structure

The artifact MeTILDA_reproduce is available at https://github.com/metilda-uw/MeTILDA_reproduce. This artifact includes a detailed

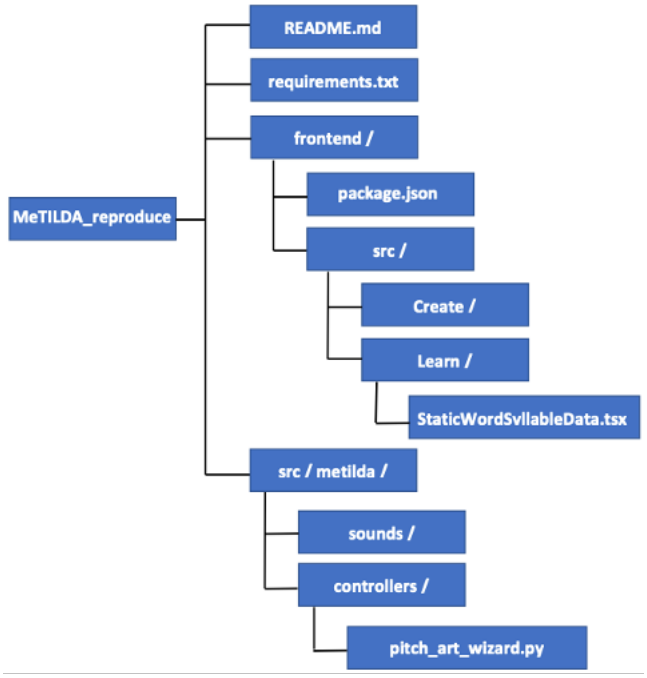

Figure 1: Code structure

README file that specifies the environmental requirements and implementation instructions.

The project is a two-tier cloud-based platform. The client tier is a Web frontend implemented using the React framework and written in Typescript. The web services tier is implemented in Python using a Flask web server. The file structure of the implementation code is shown in Figure 1. There are a significant number of files and dependencies, so we highlight the important folders and files necessary to run the system below.

README.md: This file consists of all the instructions for setting up the application.

requirements.txt: This file contains all the dependencies used in the Python Flask web server.

frontend: This directory contains the frontend code written as React web components where the application is executed. The implementation for each page in the application is contained in the sub-folders as follows:

- **frontend/package.json:** This file contains all the dependencies used by the React web components.
- **frontend/src/Create:** This directory contains the React frontend code for the "Create" page.
- **frontend/src/Learn:** This directory contains the React frontend code for the "Learn" page.
- **frontend/src/Learn/StaticWordSyllableData.tsx:** This typescript file defines a static dataset for the words available on the "Learn" page. Extending the data contained in this file will add words to the "Learn" page.

src/metilda: This directory contains the Python Flask web server application code with the following subfolders:

- **src/metilda/sounds:** This directory contains the audio files that are available for analysis on the "Create" page and audio playback on the "Learn" page. More details about the data are discussed in Section 1.3.
- **src/metilda/controllers/pitch_art_wizard.py:** This file implements the RESTful API web services that create perceptual scales, generate spectrograms, and extract phonetic features of the audio files.

1.3 Data

1.3.1 Input . The code repository includes 9 sample sound files saved to src/metilda/sounds directory. They were used for evaluation in the original paper [2]. The data used to create Pitch Art is contained in the frontend/src/Learn/StaticWordSyllableData.tsx file for each of these sample files and is used on the "Learn" page. Users can extend the system beyond these 9 samples by uploading their own .wav audio files to the sounds folder. These files can then be accessed from the "Create" page to be analyzed as described in Section 1.4.1.

1.3.2 Output. Audio syllable analysis results can be downloaded as a .json file using the "Save" button on the "Create" page. Users may use these .json files in their own applications or add them to the data structures in frontend/src/Learn/StaticWordSyllableData.tsx to extend the available words on the "Learn" page. Users may also download the Pitch Art generated during the analysis process as a .png file using the "Save Image" button.

1.4 System Features

A demo video can be accessed at https://youtu.be/U4yaQnXHXxQ or https://tinyurl.com/ICMRDemoMeTILDA, which shows how the system can be operated. More details are discussed below.

1.4.1 Create Page. As shown in Figure 2, the "Create" page provides various features to support speech analysis and Pitch Art creation.

- **Display audio spectrogram:** All the sound files can be selected from the dropdown list on the "Create" page. Once a sound file is selected, its frequency spectrogram image will be loaded, and the audio can be played using the "Play" Button.
- **Adjust pitch range:** Users may change the pitch range of a sound file to their desired values using the text fields below the dropdown list and then clicking the "Apply" button.
- **Identify syllables:** Users may select syllables from the spectrogram by clicking and dragging the mouse over the timeline. Once selected, a right-click inside the highlighted area produces a pie menu with all the analysis options, and based on the option selected the frequency is calculated and the syllable is added to the "Target Pitch".
- **Edit syllables:** Users can select the syllables from the "Target Pitch" on the "Create" page and click on the "Set Syllable" button to enter text annotations in the text field. Once saved, the text annotations can be viewed on the "Target Pitch".
- **Remove/clear syllables:** Users may select a syllable and click the "Remove" button to remove this single syllable; or use the "Clear" button to remove all the syllables.

Figure 2: The "Create Page"

- **Save/open analysis:** As discussed in Section 1.3.2, audio analysis results can be saved as a .json file using the "Save" button and view previously saved .json files by clicking the "Open" button.
- **Play/edit Pitch Art:** As the syllables are added, the Pitch Art is generated in real-time. The musical beats of the Pitch Art can be played using the "Play Tones" button. Users can adjust the look/appearance of the Pitch Art using various options provided on the left of the page.
- **Save Pitch Art:** Clicking on the "Save" button will download an image file of the Pitch Art.

1.4.2 Learn Page. The "Learn" page (see Figure 3) provides tools for practicing Blackfoot pronunciation with the following features.

- **View Pitch Art representations:** The Blackfoot words are categorized based on the number of syllables and associated with corresponding radio buttons. Clicking on the radio button options will load the Pitch Art representations of the words, which illustrate the position of accent syllables in each word.
- **Select words to learn:** Clicking on the Pitch Art image will lead to a page with all the available words in the corresponding category to learn from. Users can listen to the word pronunciation recorded by native speakers and their musical beats using the "Play" and "Play Tones" buttons.
- **Select words to learn:** Clicking on the Pitch Art image will lead to a page with all the available words in the corresponding category to learn from. Users can listen to the word pronunciation recorded by native speakers and their musical beats using the "Play" and "Play Tones" buttons.

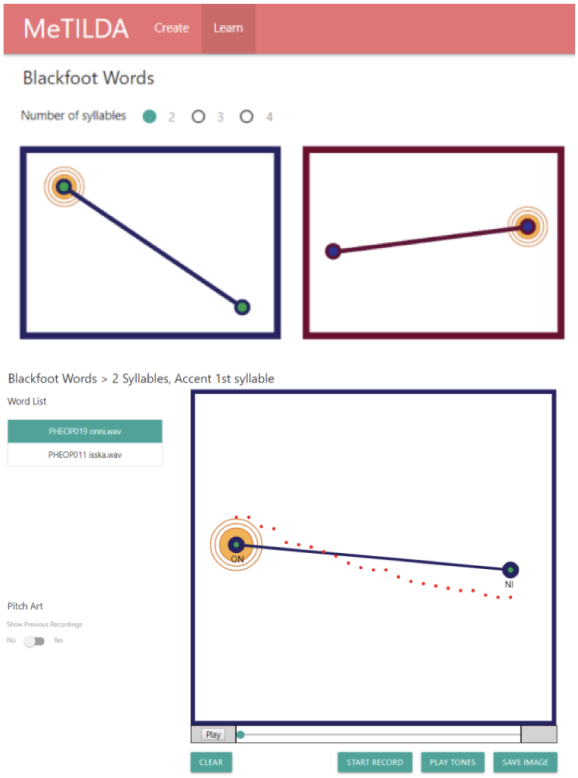

Figure 3: The "Learn Page"

- **Record pronunciation:** Users can record their own pronunciation of the word by clicking on the "Start Record" and then "Stop Record" buttons. The pitch of the recording will be represented by a red dotted line to compare with that of the native speaker. Learners can practice and record multiple times to view the changes and to improve pronunciation.
- **Manage recordings:** Users can select "Save Image" to save the pitch comparison between their recordings and the native speakers. Users may use the "Show Previous Recordings" option to view their older recordings and the "Clear" option to remove their recordings.

2 REPRODUCIBILITY EFFORTS

Installing MeTilda using the installation guide from the reproducibility GitHub repository is straightforward and easy to run. As described in Section 1.4 the MeTilda application consists of two pages, "Create" and "Learn". On the "Create" page the provided syllable files displayed their respective audio spectrogram and could be interacted with by dragging over a section using left-click. Right-clicking the section displayed the pie menu from which a syllable could be identified. Once a syllable was identified the Pitch Art was updated accordingly. All the adjustable features and buttons on the "Create" page were tested and worked as intended. On the "Learn" page multiple words with different number of syllables were tested. The main feature of this page is for the user to record their pronunciation and see how it compares with the Pitch Art. To test this

feature multiple attempts were made with varying pitches by the reviewer. The high and low pitches were displayed correctly in the Pitch Art image on the page. Each feature explained in the original work has been reproduced, with no significant issues observed.

3 CONCLUSIONS

In this paper, we presented the artifacts for the paper "MeTILDA: Platform for Melodic Transcription in Language Documentation and Application". The artifacts contain the source code, data, and documentation for system development and evaluation. Given the urgency in endangered language research, such replication efforts can help share resources and knowledge among interested individuals in academic and local communities, and enable the operation, customization, and extension of our toolsets.

ACKNOWLEDGMENTS

This work is supported by National Science Foundation (NSF BCS-2109654). We also appreciate the late Mr. Earl Old Person for his audio recording as a native speaker and learner of the Blackfoot language.

REFERENCES

[1] Naatosi I Fish. 2018. Pedagogy of pitch in L2 Blackfoot. (2018). https://scholarworks.umt.edu/utpp/193

[2] Mitchell Lee, Praveena Avula, and Min Chen. 2021. MeTILDA: Platform for Melodic Transcription in Language Documentation and Application. In *Proceedings of the 2021 International Conference on Multimedia Retrieval*. Association for Computing Machinery, New York, NY, USA, 607–610.