Cloud-based Platform for Indigenous Language Sound Education

Min Chen
University of Washington-Bothell
minchen2@uw.edu

Chris Lee
University of Washington-Bothell
leec351@uw.edu

Naatosi Fish
Blackfeet Community College/Blackfeet Nation
naatosifish@gmail.com

Mizuki Miyashita
University of Montana
mizuki.miyashita@umontana.edu

James Randall
University of Montana
james.randall@umontana.edu

Abstract

Blackfoot is challenging for English speaking instructors and learners to acquire because it exhibits unique pitch patterns. This study presents MeTILDA (Melodic Transcription in Language Documentation and Application) as a solution to teaching pitch patterns distinct from English. Specifically, we explore ways to improve data visualization through a visualized pronunciation teaching guide called Pitch Art. The working materials can be downloaded or stored in the cloud for further use and collaboration. These features are aimed to facilitate teachers in developing curriculum for learning pronunciation, and provide students with an interactive and integrative learning environment to better understand Blackfoot language and pronunciation.

1 Introduction

Blackfoot is referred to as a pitch accent language where some words can differ in meaning based only on changes in pitch (Frantz, 2017). Consider the example below where the pronunciation of the Blackfoot word apssiw has two distinct meanings based only on where the high pitch – marked with an acute symbol – is placed.

ápssiw	H*L	'it is an arrow'
apssíw	LH*	'it is a fig; snowberry'

In our previous work, a cloud-based system called MeTILDA (Melodic Transcription in Language Documentation and Application) was developed to assist in analyzing the pitch movement of individual Blackfoot words (Lee at el, 2021), and to create a visual aid called Pitch Art for learning pitch (Fish and Miyashita 2017). We are expanding on this work to support language education by improving MeTILDA's data processing, sharing, and visualization capabilities.

- Data processing: We explore working with the relative pitch differences between syllables to enhance users' understanding of pronunciation.
- Data Sharing: We are committed to community-based-research (Czaykowska-Higgins, 2009) and are extending the data storage services so users can collaboratively utilize the data-reuse capabilities of a cloud-based system when working with Pitch Art.

Proceedings of the The Seventh Workshop on the Use of Computational Methods in the Study of Endangered Languages, pages 1–6

- Data visualization: We explore ways to filter results and improve data visualization to help users identify pitch patterns in the data.

2 Related Work

2.1 Existing Linguistics Tools

Currently, existing software systems fail to address the urgent need of Indigenous language education. Language learning tools such as Babbel (n.d.) and Rosetta Stone (n.d.) are advertised as effective for language education, but they largely focus on commonly spoken languages and only have limited support for typologically distinct Indigenous languages. It is even rarer to find support for languages described by their pitch counters. Studies show that pronunciation learning technique is significantly understudied in Indigenous languages (McIvor, 2015). Pitch movements are often not explicitly represented in instructions and remain unclear to learners. On the other hand, existing linguistics tools such as Praat (Boersma and Weenink, 2013) provide essential support for linguistic research on pitch in languages. Praat is a standalone tool for acoustic analysis, providing tools to analyze sound waves and sound based spectrogram. It includes visualized pitch contours, and a feature to annotate speech sounds. However, Praat was not designed for language education because it lacks learnability and pedagogical components, and its PC-based setup does not support a collaborative learning environment that enhances education. In our previous study (Lee et al., 2021), an initial effort was done to migrate several Praat functions that are commonly used for studying pitch, such as speech synthesis, audio features extractions and a spectrum analysis. While Praat can be helpful in building collaborative linguistics training sessions for linguists and teachers, it is not intuitive for students who do not have linguistics background.

2.2 Blackfoot, Pitch Art, and MeTILDA

Previous studies on Blackfoot have identified pitch movement patterns using recordings of Blackfoot native speakers (Fish and Miyashita, 2017; Miyashita and Weber, 2020). One of these studies, conducted by a team consists of a community linguist who is also a language instructor and a formally trained linguist, became the basis for Pitch Art (Fish and Miyashita, 2017), a visual representation of pitch movement throughout a word serving as a visual aid for teachers and learners of Blackfoot (Bird and Miyashita 2018). Originally, the creation of Pitch Art involved manual processes using multiple tools such as Praat, excel, and drawing software. To lessen the burden of time-consuming procedure, MeTILDA was developed to automate the Pitch Art creation in one application (Lee et al., 2021). It also provides initial visualization to analyze and compare the speech of native speakers and that of language learners, which serves as a foundation to extend this study to assist in analyzing and teaching the Blackfoot language. The current version focuses more on teacher support, and the implementation of the system for further analysis and documentation will be conducted in the future.

3 Methods

As illustrated in Figure 1 (below), MeTILDA follows a multi-tier architecture pattern to support language education and research. It provides four major components as web services, namely Creation, Learning, and User and Content Management. All web services are deployed to the Heroku cloud platform and are made publicly available for other developers to adopt and extend the functionality in their own applications.

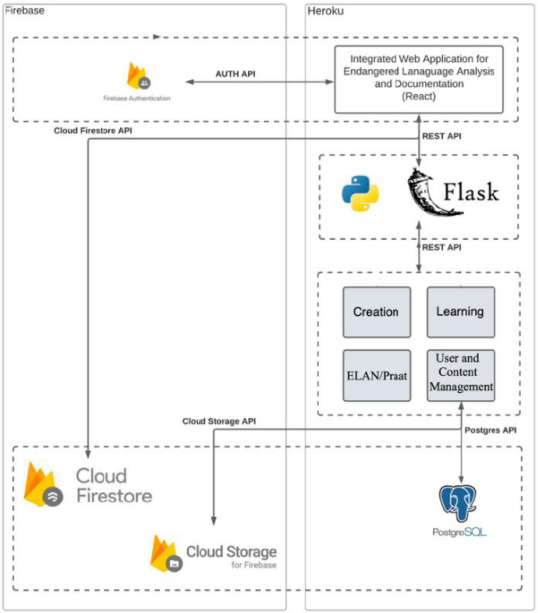
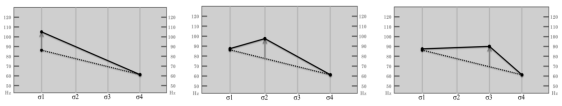


Figure 1: High-level system architecture

In addition, creating this platform as opposed to using an existing market software was drawn from the language specific community-based-research practice, which is explained in the next section where describes its Creation feature.

3.1 Creation

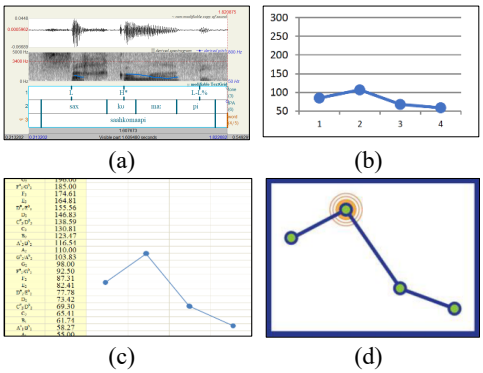
Previous research on Blackfoot has shown Blackfoot prosody in terms of unique patterns in pitch movement. (Fish and Miyashita, 2017; Miyashita and Weber, 2020). The shape of the pitch movement is predictable once the accent location, which interacts with pitch declination, is determined as shown in Figure 2. The declination starts from a mid-point in a speaker's pitch range; pitch is raised at the accented syllable which are the first syllable in (a), second syllable in (b), and the third syllable in (c) and gradually drops toward the target boundary low tone (Miyashita and Weber, 2020).



(a) (b) (c)

Figure 2: The pitch accent is on the (a) first, (b) second, and (c) third syllable.

Therefore, being able to transcribe pitch patterns becomes an important part of study in language teaching and learning techniques. Currently, Pitch Art used in some language classes (Fish and Miyashita, 2017; Bird and Miyashita, 2018). As mentioned previously, the creation of Pitch Art prior to the development of MeTILDA involved several steps and multiple tools, represented in Figure 3: (a) measuring the fundamental frequency (F0) of each vowel in Praat, (b) inputting values in an Excel spreadsheet to create a graph, (c) modifying the graph according to a perceptual scale, and (d) applying a design to the graph lines to make them more aesthetically pleasing.



(a) (b) (c) (d)

Figure 3: (a) frequency measures in Praat; (b) graph based on frequency; (c) graph on perceptual scale; (d) Pitch Art

The incorporation of a perceptual scale was determined in order to normalize the speaker-specific pitch ranges. In addition, it is challenging to identify a perceptual scale to properly reflect how pitches are auditorily perceived by human. In our previous study (Miyashita et al., 2021), we developed a MeT perceptual scale by adopting and extending the equal temperament scale in western music. The MeT scale enables the visualization of pitch data by aligning pitch data (in Hz) to a repeating series of 12 notes that form an octave. This allows users to focus on the melody or contour of the word, while disregarding the speakers' actual pitch ranges, which can vary due to age, gender, or other physical factors. We developed a feature in MeTILDA (Creation) for users to upload and process speech recordings, and to automate the creation of Pitch Art. Since the selection of parts that pitch are extracted requires knowledge in acoustic phonetics, users of the Create feature are researchers and teachers who has linguistics backgrounds. Note that the current version focuses of relative pitch movement, and it does not necessarily correspond the timing of syllable intervals.

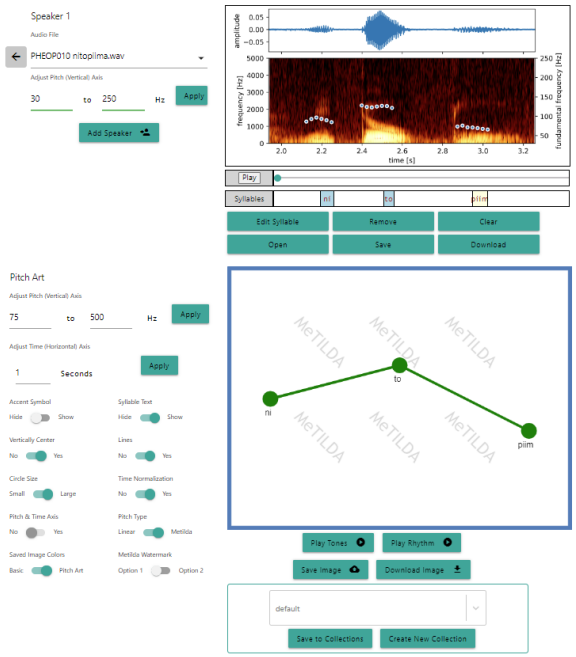


Figure 4: Creation: Audio Analysis (top) and Pitch Art (bottom)

As shown in Figure 4, the Create feature contains two main sub-components: Audio Analysis and Pitch Art. In the Audio Analysis sub-component, users can view an audio waveform and a

spectrogram of speech recordings uploaded to the cloud database. Users can either upload their own recordings or access the already-existing databases. Tools are provided for users to identify vowels and enter their orthographic symbols. We have also implemented the MeT perceptual scale in MeTILDA so pitch can be calculated and represented in both a frequency scale and a perceptual scale (Miyashita et al., 2021). Once users analyze pitch, the program automatically creates Pitch Art which then can be downloaded. Additionally, users can save Pitch Art images, measurement data, listen to the tones of the word melody, and toggle a variety of appearance options (e.g., displaying orthography, showing pitch in an F0 contour or by average, showing pitch in Hz or as transcribed to a psychoacoustic scale). The Creation component can directly contribute to the training of community linguists in acoustic phonetics and to the production of teaching materials including pronunciation guides.

3.2 Learning

The Learning component provides tools for users to learn pitch patterns, practice word pronunciations, and visualize the similarity/difference between learners' and native speakers' pronunciations as shown in Figure 5.

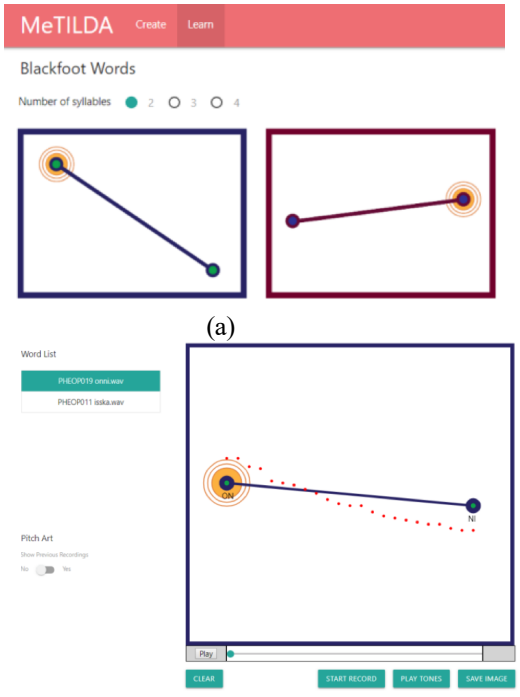


Figure 5: The Learning component

To use the Learning feature shown in Figure 5 (a), users choose a syllable pattern which is determined based on the number of syllables and the location of a pitch accent in a word (Miyashita and Weber, 2020), and listen to a native speaker's utterance. Then a Pitch Art of a sample word created based on the recording of a native speaker is shown on the screen. Then looking at the Pitch Art, users pronounce the word and record themselves, and immediately pitch tracking of the users' utterances is printed as a dotted line over the sample Pitch Art as shown in Figure 4 (b). Thus, users can compare their own pitch performance with the sample. All the recorded sounds can be saved for future access. This saving feature enables an implementation for a language course supplement, such that a teacher can students' performances.

3.3 User and Content Management

The User and the Content Management component provides the functionality to administer the MeTILDA system create or remove users, upload, move, delete files, and other administrative tasks. This component can be used to facilitate curriculum development and administer class setup/process.

Data Type	Use Case	Storage Service
Audio File (wav)	Users upload wav files for analysis.	Firebase Cloud Storage
JSON File (json)	Users save the Pitch Art data as a .json file from the Create Page.	
Image File (png)	Users save an image of the Pitch Art as .png file from the Create Page.	
Metadata about users, including user roles, and audio, image and JSON file information	Transparent to the user, this data is used by the application.	PostgreSQL
Audio Analysis Data: Words and Letter properties.	Implementation exists in the services tier to save word and letter data (e.g. time and frequency) to tables however it is not included in the front end.	
Audio Analysis Data: Words and Letter properties.	Users save the Pitch Art data as a json document from the Create Page. Introduced for the Collections feature.	Firebase Firestore

Table 1. MeTILDA Data Objects

The MeTILDA data tier makes use of multiple services in a hybrid SQL-NoSQL architecture, taking advantage of the benefits of each service for the specific data types used within the system. Table 1 describes each data type, explains its use in the system and how it maps to a service in the data storage tier. Google's Firebase Cloud Storage (Firebase, n.d.) is optimal for BLOB (binary large object) storage and is used to persist audio (wav), analysis (json), and image (png) files while Heroku

PostgreSQL (Heroku, n.d.) stores the metadata necessary for the operation of the system such as users, user roles, and metadata related to audio, analysis, and image files. In this project we introduce the use of Google Firebase Firestore, a document database optimized for real time access, to store the AudioAnalysis objects that are created as part of the Pitch Art process. To improve data re-use and the collaborative capabilities of MeTILDA, we also introduced "Collections." From the Collections page, users can create, view, rename, or delete a collection group as shown at the top of Figure 6. Once a collection group is created, users can save Pitch Art to that collection through Create component. On the Collection page, each word is displayed as a card to accommodate a thumbnail image and initial information about the PitchArt including the user supplied word, word translation, speaker name, the number of syllables in the word, and the date of creation. Clicking on the card redirects to the Learning component for review and practice.

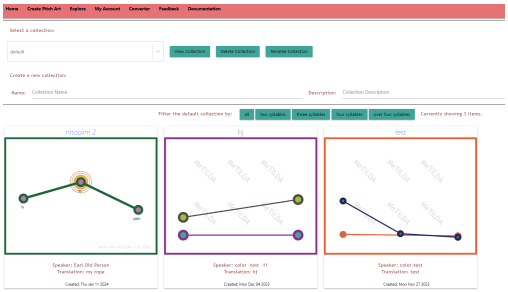


Figure 6: The Collection component

	Researcher	Teacher	Learner
Prosodic Analysis	✓	(✓)	
Upload Audio	✓	(✓)	
Create Pitch Art	✓	(✓)	
View data	✓	✓	✓
Edit Collection	✓	✓	
Play Audio	✓	✓	✓
Record Speech	✓	✓	✓
Submit Work	✓	✓	✓

Table 2. Access to Features by User Category

Users can obtain access to MeTILDA by logging in based on their user category: researcher, teacher, student, or other. These user categories were selected based on the ultimate goal for the tool to contribute to both prosody research and language teaching/learning. Different users have varied access permissions to data in the system. For example, as shown in Table 2, researchers and teachers can access all features: measurement tools, Pitch Art creation, saving, and stored data. Students do not need to create Pitch Art and have access only to play, record, and submit in the Learning component. While teachers can access all, they can optout access to the top three components that require linguistic background, and they can collaborate with researchers or other teachers who have access to all. Teachers can view their students' saved work, while students cannot see each other's submissions.

3.4 Limitations

MeTILDA is hosted on Heroku, a cloud application platform. In general, while cloud service providers implement strict security standards and industry certifications, storing sensitive and/or confidential data on external service providers requires additional measures to ensure security and privacy. Currently, MeTILDA limits the types of raw data to be processed for research activities. For example, only sound files that have acquired authorization for such use can be uploaded to MeTILDA for processing.

4 Usability Studies

While the tool is in-progress, usability surveys of MeTILDA were conducted in a linguistics class with a special topic at the University of Montana (12 students) in Fall 2022. Additionally, several language researchers via the linguist co-author's network participated in the survey. Tested components focused on those who would create Pitch Art and use edit collection features. 11 questions were chosen based on the survey research done in Lund (2001) and focused on user experience, including usefulness, ease of use, ease of learning, and satisfaction. Among them, 10 questions use the Likert format with ratings from 1 to 5 (1 being "Strongly disagree" and 5 "Strongly agree") and 1 open ended question allowing the participants to provide general feedback. A total of 25 users participated of which three were linguists, 21 were students, and 1 was a teacher. By average, the rating for each question is above 4.0 out of 5 and over half of the questions received more than 4.5 ratings, which indicates the overall effectiveness of MeTILDA in supporting the Blackfoot language education. Usability of the Learn feature has not yet been tested. More sample words for this feature need to be added before the component can be tested. However, the students in the class were shown how it works, and they

informally expressed that the component is helpful for realizing their own pitch performances.

5 Conclusions

In this paper, we presented MeTILDA, an integrated system for Indigenous language education. It supports speech data processing, analysis, visualization, and sharing via three main components: Creation, Learning, and User and Content Management. The Creation component is especially helpful for the training of community linguists in acoustic phonetics and the production of teaching materials including pronunciation guides. The Learning component helps users to learn pitch patterns, practice word pronunciations, and visualize the similarity/difference between learners' and native speakers' pronunciations using Pitch Art. User and Content Management can be used to facilitate curriculum development and administer class setup/ process. In our future work, we plan to enhance pronunciation education with other Indigenous languages.

Acknowledgments

This work is supported by National Science Foundation (NSF BCS-2109654 & NSF BCS-2109437). We appreciate the late Mr. Earl Old Person and Mr. Rod Scout for their audio recording as native speakers. We also thank students in the linguistics program at the University of Montana for participating in the survey.

References

Arnold M. Lund. 2001. Measuring usability with the USE questionnaire12. *Usability interface*, 8(2):3-6.

Babbel. n.d. Available: https://www.babbel.com/ (Accessed October 14, 2023).

Donald. Frantz. 2017. *Blackfoot Grammar*. Toronto: University of Toronto Press.

E. Czaykowska-Higgins. 2009. Research models, community engagement, and linguistic fieldwork: Reflections on working within Canadian Indigenous communities. *Language Documentation and Conservation*, 3(1):15-50.

Firebase. Available: https://firebase.google.com/ (Accessed October 14, 2023).

Florian R. Hanke. 2017. *Computer supported collaborative language documentation*. PhD. Dissertation, The University of Melbourne.

Heroku. Available: https://devcenter.heroku.com/categories/reference (Accessed October 14, 2023).

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*. Taipei Taiwan, pp. 607–610.

Mizuki Miyashita, Min Chen, James Randall, and Naatosi Fish. 2021. *Introducing the melodic transcription (MeT) scale for language documentation and application. University of Arizona Research Data Repository.* https://doi.org/10.25422/azu.data.14481825.v1.

Mizuki Miyashita and Ntalie Weber. 2020. Blackfoot Pitch Contour: An Instrumental Investigation. *Papers of the Forty-Ninth Algonquian Conference*. Eds. by Monica Macaulay and Margaret Noodin. 149-166.

Naatosi Fish and Mizuki Miyashita. 2017. *Guiding pronunciation of Blackfoot melody*. In *Honoring Our Teachers.* Eds. by J. Reyhner, J. Martin, L. Lockard & W. Sakiestewa Gilbert. Flagstaff, AZ: NAU, pp. 203-210.

Onowa McIvor. 2015. Adult Indigenous language learning in western Canada: what is holding us back? In *Living Our Languages: Papers from the 19th Stabilizing Indigenous Languages Symposium*. Eds. by Michel, K., Walton, P., Bourassa, E., Miller, J. (Eds.) pp. 37-49.

Paul Boersma and David Weenink. 2013. *Praat: doing phonetics by computer* [Computer program]" Version 5.3.51, http://www.praat.org.

Rosetta Stone. n.d. Available: https://www.rosettastone.com. (Accessed October 14, 2023).

Sonya Bird and Mizuki Miyashita. 2018. Teaching phonetics in the context of Indigenous language revitalization. In *Proceedings of ISAPh 2018 International Symposium on Applied Phonetics*. pp. 39-44.