Evaluating Indigenous language speech synthesis for education: A participatory design workshop on Ojibwe text-to-speech

Viann Sum Yat Chan, Christopher Hammerly
Department of Linguistics
University of British Columbia
Vancouver, Canada
chan.sum@northeastern.edu
chris.hammerly@ubc.ca

Abstract

This paper reports methods and results from a participatory design workshop aimed at evaluating the use of speech synthesis and text-to-speech for Ojibwe language education. Using an existing text-to-speech feature as a starting point, we worked with two groups of Ojibwe language instructors using a guided trial of the speech synthesis system and a two hour semi-structured workshop with the aim of creating a lesson plan that utilizes text-to-speech. We highlight the insights from this work, both in how to design and deliver speech synthesis systems for Indigenous language education, but also how to approach and design such a workshop to ensure a fruitful discourse.

1 Introduction

Ojibwe is a North American Indigenous language in the Algonquian family known in different regions as Anishinaabemowin, Nishnaabemwin and Ojibwemowin. It is spoken in both the US and Canada, with 25,440 speakers recorded in the 2021 Canadian Census (Statistics Canada, 2023). Colonial policies like the residential school system aimed to force assimilation through means such as reduced use of the language and separation of children from their families (Truth and Reconciliation Commission of Canada, 2015). Because of this, the Ojibwe speaker population is characterized by a high average age of L1 speakers and a parent generation who may understand the language but do not primarily speak it to their children (UNESCO, 2010).

In addition to its effects on language use within families, the lack of L1 speakers in the current parent generation also means many instructors of Ojibwe are as much learners of the language as they are teachers (Engman and Hermes, 2021). Because not all families are able to support students' language learning at home, students rely heavily on their teachers and peers in the classroom to practice the language, thus limiting their exposure to the language in other contexts and environments. Combined with the unique position of teachers as teacher-learners, the task of teaching Ojibwe poses challenges beyond what is typical of second-language learning.

One way to address this issue is through the development of synthetic text-to-speech (TTS) systems which can act as an audio supplement to existing text-based tools like verb conjugators, dictionaries, and phrasebooks (Pine et al., 2024). Currently, there are 70 Indigenous languages spoken throughout Canada, but only a handful of existing TTS systems (e.g. Harrigan et al., 2019; Pine et al., 2022; Conrad, 2020; Hammerly et al., 2023). Low-resource languages face challenges in the development of TTS due to a limited number of fluent speakers and these speakers having limited time to record data for training. Pine et al. (2024) also identifies challenges in the evaluation of Indigenous TTS systems—a small L1 population means there might not be a large enough sample to contribute to a meaningful and generalizable quantitative evaluation of the synthetic speech system. While efforts to create TTS systems have been successful, not much work has been done to investigate how language communities are using these TTS systems, and whether the intended benefits can be enjoyed.

The goal of the current study is therefore to answer the following research questions:

1. What are the strengths and limitations of our existing Ojibwe TTS feature?

2. What are teachers' priorities when approaching new tools in educational technology like TTS?

These questions address the present and the future of developing TTS for Ojibwe and other Indigenous languages. Exploring the strengths and limitations of TTS can help us troubleshoot existing problems, while understanding teachers' priorities when

using TTS in their teaching can help researchers and developers focus their improvements on the needs of the community. Observing how teachers interact with unfamiliar technology and understanding the strengths of TTS can give researchers and developers insight into what the barriers to usage are currently, and how usage of new technology can be encouraged in the future.

2 The current Ojibwe Text-to-Speech Feature

Hammerly et al. (2023) describes the development of a TTS synthesis system for Border Lakes Ojibwe that is being deployed on the *Anishinaabemodaa* web-based language learning platform produced by teams at the Seven Generations Education Institute, SayItFirst, CultureFoundry, and the University of British Columbia. Only users with "teacher" profiles are given access to the TTS feature, delivered as a standalone webpage independent of the other learning materials on the platform. The webpage (Figure 1) includes a text box for users to input text, a button to generate speech labelled "Speak!", and an audio clip once the "Speak!" button is clicked. Users can play the audio clip on the webpage or download the clip to use in different learning materials by clicking the three dots next to the audio clip to reveal a drop down menu.

As detailed in the Hammerly et al. (2023) paper, this standalone TTS feature was intended for teachers to use to generate audio files that can be sped up, slowed down or downloaded for offline use. It was also planned for teachers to be able generate their own materials and integrate the audio into games like a flashcard activity. Despite this resource being available to teachers, surveys and consultation conducted by CultureFoundry found that teachers were not using this resource, nor were they aware of it. We aim to understand why this feature has not yet seen widespread use on the platform.

3 Participatory Design and Indigenous Research Methods

Pine et al. (2024) highlighted the need for synthetic speech systems to be developed through collaboration with their respective language communities to avoid ethical issues in consent, data collection and usage. To ensure adequate community engagement and consultation in the development of our TTS tool, we seek to use participant-centred research methods to facilitate collaboration between teachers of Ojibwe, researchers and developers.

Participant-centred research methods position participants as the subject matter expert, a role traditionally held by the researcher (Zelenko et al., 2021; Flaskerud and Anderson, 1999). Participatory design (PD) or co-design is most commonly used in human-computer interaction (HCI) research as a way for users of computer technology to participate in its development, with the goal of aligning these tools with the practice and beliefs of the users (Hansen et al., 2019). It is often used to develop educational technology, inviting students, parents and teachers to contribute to the design process (Roschelle et al., 2006; Lin and Van Brummelen, 2021). While co-design focuses on creating and reporting on a tangible finished product, PD often requires the reinterpretation of the design outcomes to understand users' needs and values (Lim et al., 2008). Outcomes of PD can include intangible products like knowledge of current practice, new practices and visions for the future on top of the tangible product or prototype (Hansen et al., 2019).

PD research involving Indigenous communities place a strong emphasis on establishing a warm and welcoming environment for participants, giving participants time to build rapport, begin friendly dialogue, and get to know each other on a personal level (Parsons et al., 2016). This emphasis can be seen in researchers designating a separate workshop session for this purpose (Barcham, 2023), or spending considerable preparation time on building trust before formal data collection begins (Woodward and Marrfurra McTaggart, 2016). While participant-centred research methods have always put the spotlight on participants' voices with minimal input from the researcher (Zelenko et al., 2021), Indigenous co-design practices appear to be characterised by a disproportionately long duration of time dedicated solely to rapport-building, relative to formal data collection. Additionally, Parsons et al. (2016) recommend Indigenous co-design workshops conform to culturally appropriate ways of interacting, incorporate traditional practices in the workshop, and tie research to relevant cultural priorities.

This study is concerned with the evaluation and improvement of an EdTech tool, typical of HCI research, while also understanding the need to be respectful and sensitive of the cultural context surrounding the development of the tool. We aim to combine practices from both HCI and Indigenous

Figure 1: Screenshot of current Ojibwe text-to-speech feature on the *Anishinaabemodaa* platform

PD research to approach the design process as a vehicle for inquiry rather than simply a means to create a tangible end product.

4 Method

A single-session participatory design workshop was conducted with two groups of teachers who use the *Anishinaabemodaa* language learning platform as part of their instruction. Participants first completed a pre-workshop questionnaire and guided trial of the TTS feature on the *Anishinaabemodaa* language learning platform. The workshop involved the creation of a lesson plan that includes the use of the TTS feature and aimed to explore the capabilities and limitations of the feature through active engagement with it. Each workshop session lasted approximately two hours. All procedures were approved by the UBC Office of Research Ethics.

4.1 Participants

All five teachers who participated in the study have had contact with CultureFoundry due to their involvement with the *Anishinaabemodaa* language learning platform and were recruited through CultureFoundry's mailing list. Two workshop sessions were run, first with a group of three, then a group of two. The participants were between the ages 25 and 55 and all participants were female. Their years of experience with Ojibwe ranged from two to 51, while their years of experience teaching Ojibwe ranged from one to eight. Participants were located in Northwestern Ontario, the Greater Toronto Area and Wisconsin. All participants considered themselves learners as well as teachers of the language, and were encouraged to draw from their unique teacher-learner perspective throughout the workshop. The participants were grouped by their availability to participate in the workshop, and the group size was limited to three participants to ensure enough opportunity for everyone's ideas to be heard. Participants were paid CAD $50 per hour for their time.

4.2 Materials

4.2.1 Pre-Workshop Materials

The pre-workshop questionnaire (see Appendix A for the full set of questions) was hosted on Qualtrics and consisted of five parts: (1) a consent form, (2) a demographic questionnaire, (3) a guided trial of the TTS feature, (4) a general user-experience questionnaire and (5) a brainstorm area for participants to write down initial ideas they might have for the lesson planning activity in the workshop. The questionnaire was sent to participants prior to the workshop.

The guided trial of the TTS feature consisted of six tasks, each task prompted participants to enter a different type of text input into the TTS feature and share their first impressions. The types of text input include: (1) one word, (2) one sentence, (3) one paragraph, (4) one question, (5) sentences that convey different emotions, and (6) any other text input they would like to try.

Each task was structured in the same way: the participants were first prompted to try entering one type of text into the TTS feature, then, they were to type out the text input they chose, listen to the synthetic speech output generated by the TTS feature, and rate how accurate, natural and contextually appropriate the speech sounded on four-point Likert scales. In the final two tasks, participants had the option to try out additional phrases, sentences or paragraphs and report their thoughts in more detail

in an open-ended text box. The purpose of this guided trial was to ensure participants had interacted with the TTS feature in a meaningful way, and had an opportunity to discover the capabilities and limitations of the TTS feature on their own terms prior to involvement from other participants and the researchers.

4.2.2 Workshop Materials

The workshop was hosted on Zoom and the Whiteboard feature was used for collaboration between participants. The Whiteboard was set up prior to the session with four different areas (Figure 2), moving from one area to the next as the workshop progresses. The first is a brainstorm area framed by a white square where participants can add "sticky notes" with their ideas on how to incorporate the TTS feature into a lesson or resource. Prior to the workshop, sticky notes with ideas that were previously suggested in the brainstorm portion of the pre-workshop activity were placed onto the brainstorm area on the Whiteboard. The second is the sorting area which included three rectangles labelled "Let's discuss this!", "Maybe discuss these if we have time" and "Save for another day". Participants were expected to move their sticky notes and sort each idea into one of these three boxes. The third area included three examples of built-in templates that can be used for lesson planning. There are many templates to choose from on Zoom Whiteboard, this sample template area was meant to give suggestions but not limit what participants eventually chose to use in lesson planning. The fourth area was the lesson planning area, used to create the lesson plan or resource together to reach a final product.

To guide participants through the introduction and discussions, a PowerPoint presentation with a progress bar was created. The same progress bar was included on the Zoom Whiteboard.

4.3 Workshop Design

The two-hour long workshop sessions were planned as described below, but we were flexible with our approach and did not follow it strictly. Changes in plans are addressed in Section 8, and full details of workshop plans, design, goals and time management are included in Appendix B.

Each workshop started with a Welcome presentation and self-introduction activity to help participants warm up and build rapport. This was followed by a brainstorm task for participants to share their ideas on how to incorporate the TTS feature into a lesson plan. Sorting tasks were planned for participants to parse through these ideas but these tasks were skipped, and participants directly identified one idea to develop further. This led into the design of a full lesson plan from the idea that was chosen and wrapped up with a workshop debrief and reflection.

5 Workshop Products

Through workshop discussions and activities, participants in the two workshop sessions created the following lesson plans to incorporate the TTS feature into their teaching.

Group 1 designed a make-your-own phrasebook activity where students would create their own customizable digital phrasebook. Teachers would model to students how to add new phrases they come across in daily life to the digital phrasebook in text and audio form and encourage usage of this phrasebook outside the classroom. The audio clip would be created with the TTS feature. The full lesson plan and additional ideas from Group 1 are included in Appendix C.

Group 2 designed a make-your-own flashcards activity. Teachers would model to students how to create digital flashcards. Students are meant to listen to audio clips of target words or phrases repeatedly and practice their pronunciation at home. When ready, they can record themselves saying these words and phrases, and embed the audio clips onto the flashcards. This activity can double as an oral assessment. The full lesson plan and additional ideas from Group 2 are included in Appendix D.

6 Synthetic Voice Quality

The pre-workshop questionnaire revealed a number of interesting results. There is a consensus between participants that while the TTS feature does not produce speech that sounds contextually appropriate the pronunciation of specific words and phrases are accurate. Pre-workshop questions on whether participants believed the synthetic speech sounded accurate received 11 responses rated "Strongly Agree", 8 rated "Agree" and 1 rated "Disagree" (See Appendix A for full results). Participant 4 further highlighted in a questionnaire response: "I tried the glottal sounds and a few other different sounds we have that are unique (different from English) [...] and all were pronounced correctly." In regards to the TTS system's ability to differ-

Figure 2: Zoom Whiteboard set up including a progress bar, brainstorm area, sorting area and lesson planning area

entiate between similar sounds, Participant 4 also suggested in a questionnaire response that the TTS feature would be a good tool to demonstrate how misspelling leads to a change in morphemes and results in words that look similar but are different in meaning. They give the example of the first person suffix *-yaan* versus the second person suffix *-yan*, which differ only in vowel length, so are frequently confused. This accuracy makes it possible for students to use the TTS feature as a secondary resource for speaking and listening practice. Students need as many reference points for the language as they can get and it is important for them to "hear a voice other than [their teachers']" (Participant 2). However, because the synthetic speech lacks natural rhythm and tone modulation (Participant 5 on questionnaire), the feature is better used for pronunciation practice than conversation practice.

7 Teachers' Priorities

We identify four priorities based on direct feedback on the TTS feature and language learning platform provided by participants on the pre-workshop Qualtrics questionnaire, workshop discussions, participants' approach to the lesson planning task and additional responses to personal reflection questions.

7.1 Representation

Participants appreciated that the TTS feature and the synthetic speech used across the online learning platform 'can allow students to hear the language from a voice other than [theirs]' (Participant 2) because a lot of their students come from families who do not speak the language at home, remarking that 'even though it is synthetic it does sound spot on' (Participant 4). However, when asked how to make the feature more culturally relevant to its potential users, participants across both workshops suggested the inclusion of different voice options, as there is currently only one voice of a middle-aged male behind the synthetic speech output. Participants highlighted the importance of having a female voice on the feature:

- 'It is important for kids to hear female voices and know that men aren't the only speakers [of Ojibwe], there are great female speakers out there as well.' (P4)
- 'There might be some trauma with men, so if they have a voice they felt comfortable with, that might be [a good] option as well.' (P5)

Choice and autonomy are key to recovery from trauma related to gender-based violence (Elliott et al., 2005), and having the option of a female synthetic voice would support that.

The importance of having a younger voice on the feature was also highlighted:

- 'It would be amazing for young folks to hear the language spoken accurately by a young sounding speaker, not necessarily culturally relevant but definitely more relevant to young people.' (P3)
- 'My kids know on fun days I play TV shows dubbed over in Anishinaabemowin like Spongebob or Scooby Doo, and they always think it's hilarious that the voices are much older than the characters they are portraying. In high school we talk about why that is, and it's obviously a serious concern that so many of our fluent speakers are getting so old.' (P2)
- 'A kid voice would be more engaging, especially since there is only one voice on the language learning platform' (P2)

Finally, participants suggested synthetic speech as a means to preserve the voices of elders, saying, 'We're losing our elders and we will lose their

voices as well' (P5). Participant 5 gave the example of a feature on the Ojibwe People's Dictionary which allows you to choose between voices of different elders when playing recordings of words by clicking on the elder's initials, as a great way to add more voice options and pay tribute to important members of the community.

7.2 Accessibility

Participants' concern with access was three-fold: the TTS feature should be more accessible on the app, the interface should include accessible language and user-friendly buttons, and there was general concern for access to technology in rural areas.

Currently, the TTS feature is only made available to teachers and it takes four clicks to reach the interface from the home page. Furthermore, awareness of the feature among teachers is limited. Participants expressed that it was through this workshop that they first heard of the feature. Just knowing that the feature is available and understanding what it is for would be huge steps in increasing access and usage. Additionally, specific parts of the TTS feature like the download, slow down and speed up functions are hard to locate. Participants appreciated these functions when told about them, but crucially needed to be told explicitly about their existence and where to access them.

When asked to complete a guided TTS trial in their own time prior to the workshop, participants reflected that there was a learning curve and using the TTS feature was not an intuitive experience. There is a button on the TTS interface labelled "Speak!" under the text box to indicate that the user is telling the TTS program to speak, but several participants thought this was an instruction for them to speak to the TTS feature and record their own voice. Participant 2 suggested this label should be replaced with the phrase "Generate Speech" which is more straightforward and tells users exactly what the button does. In trying to avoid technical language or jargon to make tools more user-friendly, the actual meaning of the instruction might be lost and have the opposite effect to the accessibility that word choice was intended to achieve.

Brief interactions with the TTS feature before the workshop already revealed several barriers to access. Participants who worked in rural school districts brought up barriers to access in terms of internet connection and access to a device at home as an additional hurdle. This makes it difficult for students to access the benefits of using the TTS feature at home, such as aiding in independent study and practicing the language in private. Along with the lack of exposure to devices at home comes with an unfamiliarity towards educational technology in general, meaning the learning curve for these students would be steeper than those who have been using all sorts of technology in their learning across different subjects. Certain rural school districts limit access to the online platform to only high school students because the technology is too hard to use, thus widening the gap in access to Ojibwe language learning resources between students in rural and urban school districts.

Because access to internet is an issue, Participants 4 and 5 particularly expressed their appreciation for the download function of the TTS feature, as it can be used to create offline multimodal resources.

7.3 Encouraging Language Usage

Encouraging usage of the Ojibwe language itself as well as the resources for language learning emerged as a priority for teachers. Participant 3 approached her teaching based on the idea that "The only wrong way to speak your language is to not speak it at all." This means getting students to engage with the language as much as possible regardless of how accurate or "good" they are. Participants liked that the TTS feature offers students a chance to practice their pronunciation independently at home by listening to the audio clips and copying the sounds. This is especially key as some students' families do not speak the language, and they rely on their teacher and lesson time to practice interactive language-use.

Another barrier to increasing language-use is students getting self-conscious. Participant 5 offers students the option to take their oral assessments or activities to a private room to complete independently, which does help students feel more comfortable, but might not be conducive to the maximized language exposure needed for effective language acquisition (Matusevych et al., 2017). The TTS feature can help these students gain exposure to the language without opening themselves up to the social anxiety of speaking to a figure of authority like a teacher, elder, or older family member.

In the lesson plans created by both groups, the first part of the lessons involved the teacher directly modelling how to use the technology. This suggests

the first barrier teachers and students face when being introduced to new tools is always the simple question of "How do I use this?" Following initial instruction, participants across both workshops had a plan for encouraging habitual usage of the resource built into their lesson plans. The participants in the first workshop session included a plan to add their phrasebook to the class' daily routine. Participant 3 suggested incorporating this phrasebook into her class' existing word-of-the-day routine—asking students to record these phrases and words in their phrasebook, while also reminding them to use the phrasebook throughout the day. Both the phrasebook and flashcard activities were designed in a way that allows students to continuously add to the resources created, with the goal of helping students build the habit of language learning in their day-to-day lives, outside of school, creating "a living document of [the students'] learning" (P2). These considerations are in line with Indigenous views that learning is "a life-long, self-directed process of experiencing, processing and reshaping existing knowledge," (English, 2008), without the distinction between adult-learning and K-12 education typical of Western conventions. The priority of encouraging language-use is perhaps a reflection of cultural values held by teachers of Ojibwe, as well as a desire to document and revitalize the language.

Encouraging language usage means involving families and community members so students can practice the language in different contexts. Both workshop products included an element of allowing students to take their work home and show their parents as a way to help parents learn the language alongside their children. The phrasebook or flashcards created can be as much a resource for parents as it is for students, and students are encouraged to continually add to these resources outside of school, which can be a bonding activity for families.

Participant 2 also mentioned how other teachers in her school who do not speak or teach Ojibwe have expressed the desire to learn a few words in Ojibwe to use with the students so they can hear the language from more people and in more contexts. Participant 2 suggested that the TTS feature would be a great resource for these teachers to practice and look up the pronunciation of certain words they had forgotten, making it easier for them to be a part of the community. This benefit can also be extended to teacher-learners of Ojibwe who are not completely fluent in the language.

7.4 Inclusion

The inclusive education framework Universal Design for Learning (UDL; CAST 2024) encourages teachers to create multimodal resources that offer multiple means of representation so students with a range of needs can access the same lesson in different ways. For instance, an audio clip next to a chunk of text would help students who have difficulty reading understand the content and having both modalities would be helpful for all L2 learners regardless of their needs. Participant 1 said teachers are "constantly recording [themselves] to create materials" for their classes, and Participant 5 was delighted to find out about the download button, commenting, "I know what I'll be playing with this evening!" The download function of the TTS feature makes something that teachers were already doing more convenient, so it is easy for them to integrate this standalone feature into their existing teaching practices.

Multimodality was heavily considered in the design of the first group's lesson plan, not only in the inclusion of both audio and text, but also in adding cross-curricular elements like having students create a customized background for their phrasebook so the phrasebook feels like their own or a themed background to match the content. Participant 2 suggested that a student interested in basketball phrases can decorate their page with basketball drawings. Participants prioritized offering students a comprehensive learning experience that does not stop at the text and the language itself, and can benefit a range of students who might prefer to learn in different ways.

UDL (CAST, 2024) also calls for multiple means of expression, meaning teachers should offer different assessment pathways for the same content to cater to diverse needs. Participants in the second workshop session highlighted challenges faced by teachers in providing accommodations and modifications for students in a school subject lacking in standard resources and practices, especially since creating custom materials adds to teachers' workload. Efforts to differentiate are often covert, designed so students are unaware of it. For one module, Participant 4 offered three different modes of assessment, one of which was a Kahoot quiz that appeared to be a lighthearted and interactive activity for the whole class, but actually assessed students who struggled with plain text. Such considerations were apparent in this group's lesson

plan, which involved students creating multimodal flashcards with text and audio clips of students' own voice recordings made after practicing pronunciation with the TTS feature. This activity offers teachers the opportunity to assess students orally, while also being a hands-on activity students can enjoy without feeling like they are being assessed.

There is a need to differentiate because a number of students struggle with language learning, even with English. Participants 4 and 5 raised concern about Ojibwe being the harder language, and having to learn it as L2 when students' L1 English abilities are not up to grade level is particularly challenging. For these students, the greatest barrier to using the TTS feature is in the feature's adherence to the standard Double Vowel orthography, which they report is not taught in certain school districts. This indicates a broader problem of literacy in Ojibwe and English, rather than an issue with the TTS feature design per se. However, the TTS itself can be a useful tool for those struggling to read a given text, since students could use the TTS feature to listen while reading along to a passage to aid in their comprehension. Participant 5 wrote in a questionnaire response, "I would think many students who are not strong in English language will have difficulty as they would also not have a strong grasp of Ojibwe words. The words need to be in front of them to be able to type it in properly and be able to identify the word. Without having the properly spelled words in front of you, if you misspell the word, [the TTS feature] does not correct it." Moving forward, one direction for our work could be to integrate a spell-checking mechanism into the TTS input, which could correct deviations from the standard orthography. We could also explore the possibility of expanding to other writing systems in the language such as syllabics.

8 Lessons Learned

8.1 Building trust and rapport is as much a priority as meeting the aims of the study

The emphasis on rapport and trust building in Indigenous participatory design research is reflected in our flexible approach to the workshop design—not intentionally allotting too much time to unstructured chatting in our plans, but allowing conversations to run as long as participants felt comfortable doing so. In line with practices in other EdTech participatory design workshops (Lin and Van Brummelen, 2021), the first workshop session we held started with a formal welcome presentation where the goal of our research was discussed, so participants and researchers were on the same page. Participants listened with their microphones muted throughout the presentation, until they were prompted to introduce themselves. There were questions on a PowerPoint presentation slide providing suggestions to guide their introductions, and the facilitator introduced herself with those same questions first to give participants time to prepare. These questions included a mix of personal, professional and lighthearted, fun questions. For a more reserved group like this one, having these questions might help participants warm up without overwhelming them.

In the second workshop session, despite never having met, the participants dove into an open discussion on the challenges faced by teachers in their communities before the formal welcome presentation. Insights shared in this unstructured time were incredibly valuable, and we believe that unprompted comments gave the best representation teachers' priorities. This went on for over 30 minutes before the workshop started as planned.

The rapport-building portions of these two sessions of the workshop went differently, yet both were beneficial for their respective groups and fit the personalities of the participants. It is important for researchers to anticipate and hold space for both possibilities. Helping participants balance open-endedness and freedom to speak with the pressure of having to come up with new ideas on the spot is the facilitator's job during the workshop and should be heavily considered in workshop plans as well.

In addition to prioritizing rapport-building, Participant 1 reflected that simple yet explicit mention that "This is a safe space," was already helpful in making her feel more comfortable. Encouragement and positive feedback throughout the workshop can also contribute to this welcoming environment, but feedback should be kept non-specific so as to not influence participants' opinions.

The benefits of rapport-building and interactive workshops can be seen in our study. CultureFoundry regularly solicits feedback on the *Anishinaabemodaa* language-learning platform, but this workshop process has helped teachers generate new ideas on how the platform can be improved. The participants in this TTS-focused study had lots of additional ideas for the language-learning platform in general, which suggests the collaborative inquiry

done in this study can be further extended to other EdTech tools and platforms in the future.

8.2 Role reversal in researcher-participant dynamics

The role reversal between researchers and participants in participant-centered research design means researchers can afford to be more flexible in their study design.

Researchers are trained to be precise in their methods, focusing on sensitivity and validity in their experimental design in order to elicit a meaningful outcome in data analysis (Lipsey and Hurley, 2009). Along with this mindset comes research anxiety, referring to how researchers can feel pressured throughout the research process to design methods, collect data and analyze results in a way that is publishable (Cooper et al., 2023).

In the case of the current study, we had discussion questions prepared for the workshop debrief and reflection as a way to guide the conversation towards being relevant to our research questions. However, in the second workshop, the participants started speaking freely, independent of any input from the facilitator, and had already addressed the research questions before the workshop formally started. In addition, we had designed a sorting task to prompt participants to consider and explain their decision-making in greater depth. Both groups opted not to participate in the sorting task and moved straight to choosing an idea to further develop. Because of how willing participants were to share their thoughts and expertise without prompting, the sorting task likely would not have added any more depth to the conversation. In retrospect, the inclusion of this task was intended as a way for researchers to feel confident in the richness of the data collected. Unlike most scientific research, changing the methods and being flexible did not impact the quality of the data in this study. Teachers who are passionate about their work will tell you what their priorities are without prompting.

In research where the participants are consulted for their expertise, researchers should approach the design in an exploratory manner, which might go against their training but will ultimately be rewarding. While helping participants feel as though they can trust the researchers is key to effective collaboration, here, we see how that trust can go both ways.

9 Conclusion

The goal of this study was to understand the priorities of teacher-learners of Ojibwe when approaching new tools in EdTech like this text-to-speech feature, and this was achieved through the pre-workshop activity, the workshop itself, lesson plans generated, and post-workshop interview. Participants cared about how well the synthetic speech represented their community, how easily the feature can be accessed, how they can encourage their students to use the language and the tools available, as well as how the TTS feature can be used to aid in inclusion.

We were also interested in the strengths and limitations of our existing Ojibwe TTS feature for the purpose of improving it for its users. Feedback provided by teacher-learner participants exposed gaps between what developers of the online platform and teachers understand as "accessible". This highlights where more work needs to be done in consultation of teachers or users of new digital tools in order to better serve the community. Additionally, an area in which representation is lacking on the online platform was revealed in our use of a single synthetic voice. However, there is great potential in how the TTS feature can be improved and used. It is possible to use synthetic speech as a document of the different voices in the community. The feature can also be useful in its ability to help teachers create multimodal resources conveniently and involve more of the community in supporting students who are developing their language skills.

Limitations

As is common for research on low-resource languages, there was a limited pool of possible participants who were available to participate in this research and this was reflected in our small sample size of five. This workshop was lengthy and required a time commitment of at least three hours as well as access to a computer and internet connection. This proved difficult for some of our participants but our workshop plans were kept open and flexible in anticipation of these challenges.

Similar workshops in the past often involved multiple sessions, with one independent rapport-building session. For the purpose of this limited project focusing on a simple TTS feature, the choice to do one session was appropriate as we were mindful of challenges in recruitment, but participants did have much to share and the session

could have run for longer if not for time constraints.

The choice to conduct the workshop on Zoom was due to logistical reasons, with participants spread out across Ontario and one participant in Wisconsin, while the researchers were based in Vancouver, BC. As much as possible, this kind of research should be conducted in-person as it would be beneficial for rapport-building and communication. It would also be much easier to run a lesson plan creation workshop with sticky notes, pen, paper and other physical materials than a blank digital space like the Zoom Whiteboard which participants were unfamiliar with and found difficult to use.

Acknowledgements

This work would not be possible without the teachers who so generously shared their time, insight, professional expertise and personal experiences with us, and continue to do so in their work every day. We thank Nozomi Nagashima at CultureFoundry for helping with recruitment and liaising with teachers. We also thank Hope and Danielle for their participation in the pilot session of the workshop. Special thanks to Dongwook Yoon for help in the ideation process for our research methods, for comments and feedback throughout, and to Jian Zhu for helpful feedback as well. Finally, we thank Jason Jones for lending his voice to the Ojibwe text-to-speech feature and developers at CultureFoundry for creating and maintaining this tool.

References

Manuhuia Barcham. 2023. Towards a radically inclusive design–indigenous story-telling as codesign methodology. *CoDesign*, 19(1):1–13.

CAST. 2024. Universal Design for Learning guidelines version 3.0.

Michael Conrad. 2020. Tacotron2 and Cherokee TTS.

Katelyn M Cooper, Sarah L Eddy, and Sara E Brownell. 2023. Research anxiety predicts undergraduates' intentions to pursue scientific research careers. *CBE—Life Sciences Education*, 22(1):ar11.

Denise E Elliott, Paula Bjelajac, Roger D Fallot, Laurie S Markoff, and Beth Glover Reed. 2005. Trauma-informed or trauma-denied: Principles and implementation of trauma-informed services for women. *Journal of community psychology*, 33(4):461–477.

Mel M Engman and Mary Hermes. 2021. Land as interlocutor: A study of Ojibwe learner language in interaction on and with naturally occurring 'materials'. *The Modern Language Journal*, 105(S1):86–105.

Jacquelyn H Flaskerud and Nancy Anderson. 1999. Disseminating the results of participant-focused research. *Journal of Transcultural Nursing*, 10(4):340–349.

Christopher Hammerly, Sonja Fougère, Giancarlo Sierra, Scott Parkhill, Harrison Porteous, and Chad Quinn. 2023. A text-to-speech synthesis system for Border Lakes Ojibwe. In *Proceedings of the Sixth Workshop on the Use of Computational Methods in the Study of Endangered Languages*, pages 60–65.

Nicolai Brodersen Hansen, Christian Dindler, Kim Halskov, Ole Sejer Iversen, Claus Bossen, Ditte Amund Basballe, and Ben Schouten. 2019. How participatory design works: mechanisms and effects. In *Proceedings of the 31st Australian conference on human-computer-interaction*, pages 30–41.

Atticus Harrigan, Timothy Mills, and Antti Arppe. 2019. A preliminary Plains Cree speech synthesizer. In *Proceedings of the Workshop on Computational Methods for Endangered Languages*, volume 1.

Youn-Kyung Lim, Erik Stolterman, and Josh Tenenberg. 2008. The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 15(2):1–27.

Phoebe Lin and Jessica Van Brummelen. 2021. Engaging teachers to co-design integrated ai curriculum for K-12 classrooms. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pages 1–12.

Mark W Lipsey and Sean M Hurley. 2009. Design sensitivity: statistical power for applied experimental research. In *The SAGE handbook of applied social research methods*, pages 44–76. SAGE Publications, Inc.

Yevgen Matusevych, Afra Alishahi, and Ad Backus. 2017. The impact of first and second language exposure on learning second language constructions. *Bilingualism: Language and Cognition*, 20(1):128–149.

Meg Parsons, Karen Fisher, and Johanna Nalau. 2016. Alternative approaches to co-design: insights from indigenous/academic research collaborations. *Current Opinion in Environmental Sustainability*, 20:99–105.

Aidan Pine, Erica Cooper, David Guzmán, Eric Joanis, Anna Kazantseva, Ross Krekoski, Roland Kuhn, Samuel Larkin, Patrick Littell, Delaney Lothian, Akwiratékha Martin, Koren Richmond, Marc Tessier, Cassia Valentini-Botinhao, Dan Wells, and Junichi Yamagishi. 2024. Speech generation for indigenous language education. *Computer Speech Language*.

Aidan Pine, Dan Wells, Nathan Brinklow, Patrick Littell, and Korin Richmond. 2022. Requirements and

motivations of low-resource speech synthesis for language revitalization. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7346–7359.

Jeremy Roschelle, William Penuel, and Nicole Shechtman. 2006. Co-design of innovations with teachers: Definition and dynamics. In *Proceedings of ICLS 2006*, volume 2. International Society of the Learning Sciences.

Statistics Canada. 2023. Indigenous languages in Canada, 2021.

Truth and Reconciliation Commission of Canada. 2015. *Honouring the Truth, Reconciling for the Future: Summary of the Final Report of the Truth and Reconciliation Commission of Canada*. House of Commons.

UNESCO. 2010. Atlas of the world's languages in danger. United Nations Education, Scientific and Cultural Organization.

Emma Woodward and Patricia Marrfurra McTaggart. 2016. Transforming cross-cultural water research through trust, participation and place. *Geographical Research*, 54(2):129–142.

Oksana Zelenko, Rafael Gomez, and Nick Kelly. 2021. Research co-design: meaningful collaboration in research. In *How to Be a design academic*, pages 227–244. CRC Press.

A Qualtrics Questionnaire

Welcome to our research study evaluating the quality of an Ojibwe text-to-speech feature. This questionnaire will include some demographic questions, followed by a guided trial of the text-to-speech feature on the *Anishinaabemodaa - Waking Up Ojibwe* language learning platform. Be sure have this questionnaire and the text-to-speech feature open on your screen at the same time so you can follow along. At the end of the questionnaire, you will be asked to give some ideas on how to incorporate the text-to-speech feature into your instruction, or some ways you can use it as a learner. The questionnaire should not take more than an hour. Thank you for your time!

A.1 Demographic Questions

1. What is your current age?
2. At what age did you start learning Ojibwe?
3. How long have you been teaching Ojibwe?
4. Would you consider yourself a learner of the language alongside being an instructor? (Yes/No)

Read each "I can..." statement and think about which answer best describes where you are in your usage of Ojibwe (Likert Scale: "Not Yet", "Rarely", "Sometimes", "Mostly", "Always")

1. I can sound out individual words
2. I can accurately spell individual words
3. I can initiate a conversation and stay on topic
4. I can recognize individual words when listening to elders speak
5. I can understand whole sentences when listening to elders speak
6. I can understand what elders say and I am able to identify the main idea

A.2 Text-to-Speech Guided Trial

The following guided trial of the text-to-speech feature will involve entering five different kinds of text into the text-to-speech system, and evaluating the synthetic speech output. You will be asked to record what you entered into the system and share your impressions of the output. The questions will ask you to input one of each type of text, but you are encouraged to experiment with more than one word, phrase or sentence; be sure to record all of them in the text box. You can use words, phrases and paragraphs from textbooks or any existing media, but feel free to come up with your own ideas and other kinds of text input we have not listed. There will be an opportunity for you to record anything else you have tried at the end of the guided trial.

Question 1a: Try entering one word into the text-to-speech system, write down what you entered in the text box below: [text box]
Question 1b: To what extent do you agree with the following (Likert Scale: Strongly Disagree, Disagree, Agree, Strongly Agree, N/A):

- The word was sounded out accurately
- The tone of voice was contextually appropriate

Question 2a: Try entering one sentence with at least three words into the text-to-speech system, write down what you entered in the text box below: [text box]

Question 2b: To what extent do you agree with the following (Likert Scale: Strongly Disagree, Disagree, Agree, Strongly Agree, N/A):

- The words were sounded out accurately
- The tone of voice was contextually appropriate
- The transitions between words sounded natural

Question 3a: Try entering one paragraph with at least three sentences into the text-to-speech system, write down what you entered in the text box below: [text box]
Question 3b: To what extent do you agree with the following (Likert Scale: Strongly Disagree, Disagree, Agree, Strongly Agree, N/A):

- The words were sounded out accurately
- The tone of voice was contextually appropriate
- The transitions between words sounded natural
- The transitions between sentences sounded natural

Question 4a: Try entering one question into the text-to-speech system, include a question mark in your input, write down what you entered in the text box below: [text box]
Question 4b: To what extent do you agree with the following (Likert Scale: Strongly Disagree, Disagree, Agree, Strongly Agree, N/A):

- The words were sounded out accurately
- The tone of voice was contextually appropriate
- The transitions between words sounded natural

Question 5a: Try entering sentences that convey different emotions into the text-to-speech system, write down all sentences you entered in the text box below: [text box]
Question 5b: To what extent do you agree with the following (Likert Scale: Strongly Disagree, Disagree, Agree, Strongly Agree, N/A):

- The words were sounded out accurately
- The tone of voice was contextually appropriate
- The transitions between words sounded natural

Question 6: Please share anything you found interesting from trying out the different sentences. Did the results meet your expectations? Was there anything you found surprising? [text box]
Question 7a: Feel free to experiment with the text-to-speech feature and come up with new ideas to enter into the system. Write down what you entered in the text box below: [text box]
Question 7b: Please share any interesting observations or insights from your additional experiments: [text box]

A.3 User Experience Questions

To what extent do you agree with the following statements (Likert Scale: Strongly Disagree, Somewhat Disagree, Somewhat Agree, Strongly Agree)?

1. The text-to-speech feature is easily accessible on the platform
2. The text-to-speech feature is easy for a new user to navigate with no prior knowledge of the feature

3. The text-to-speech feature is easy to use
4. The text-to-speech feature is able to generate synthetic speech output in a timely manner
5. I can hear the synthetic speech output clearly
6. I can understand the synthetic speech output clearly
7. Organization of information on the screen is clear and easy to follow
8. The text-to-speech feature can be useful for individuals seeking to improve their general fluency in Ojibwe

Please share any other first impressions from interacting with the text-to-speech feature that you would like to highlight. Were there any results that were unexpected or surprising? [text box]

A.4 Lesson Plan Ideas

We are interested in new and innovative ways to use the text-to-speech feature. Please use the following space to write down between three and five ideas you have on how to incorporate the text-to-speech feature into an existing or new lesson activity, OR how you might use this feature as a learner of the language. You will be expected to share these ideas with other teacher participants during the co-design workshop. [text box]

A.5 Guided Trial Response Summary Table

	Question	Strongly Agree	Agree	Disagree	Strongly Disagree
Enter one word	The word was sounded out accurately	3	1	0	0
	The tone of voice was contextually appropriate	2	2	0	0
Enter one sentence	The words were sounded out accurately	1	3	0	0
	The tone of voice was contextually appropriate	1	2	1	0
	The transitions between words sounded natural	0	4	0	0
Enter one paragraph	The words were sounded out accurately	1	2	1	0
	The tone of voice was contextually appropriate	2	2	0	0
	The transitions between words sounded natural	1	2	1	0
	The transitions between sentences sounded natural	1	3	0	0
Enter one question	The words were sounded out accurately	3	1	0	0
	The tone of voice was contextually appropriate	1	1	2	0
	The transitions between words sounded natural	2	2	0	0
Enter sentences that convey different emotions	The words were sounded out accurately	3	1	0	0
	The tone of voice was contextually appropriate	0	1	3	0
	The transitions between words sounded natural	2	2	0	0
Total	The words were sounded out accurately	11	8	1	0
	The tone of voice was contextually appropriate	6	8	6	0
	The transitions between words sounded natural	5	10	1	0

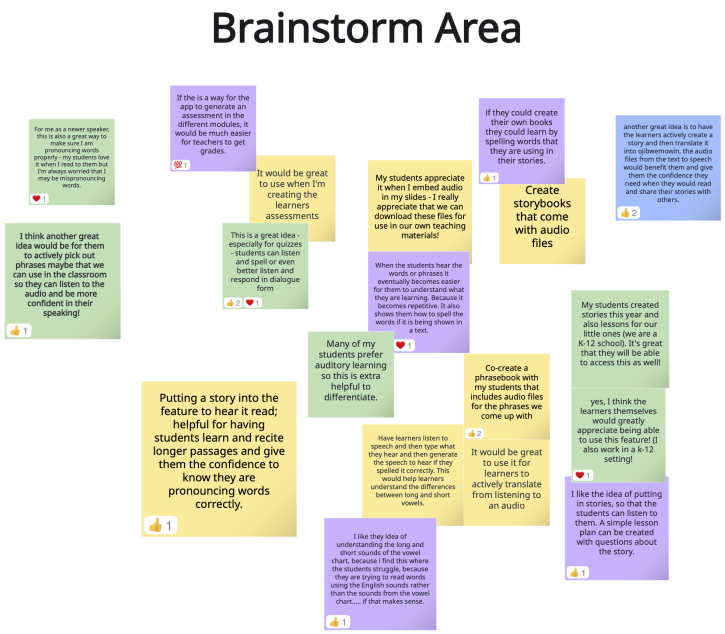
A.6 User Experience Questions Response Summary Table

Question	Strongly Agree	Somewhat Agree	Somewhat Disagree	Strongly Disagree
The text-to-speech feature is easily accessible on the platform	2	0	3	0
The TTS feature is easy for users to navigate without prior knowledge	2	0	3	0
The TTS feature is easy to use	3	0	2	0
The text-to-speech feature is able to generate synthetic speech output in a timely manner	4	0	1	0
I can hear the synthetic speech output clearly	4	0	0	1
I can understand the synthetic speech output clearly	4	0	1	0
Organization of information on the screen is clear and easy to follow	3	0	1	1
The text-to-speech feature can be useful for individuals seeking to improve their general fluency in Ojibwe	5	0	0	0

B Workshop Design Details

Activity	Duration	Goal	Additional details
Welcome and introductions	15-20 minutes	Warm up and build rapport	Land acknowledgment; explanation of the research and what to expect during workshop; encourage participants to think out loud and talk through decision making. Participants introduced themselves in Ojibwe, sharing spirit name and connection with the language.
Secondary brainstorm	10-15 minutes	Additional opportunity to share ideas, perhaps ones inspired by others	Facilitator modelled how to add sticky notes to Whiteboard. Participants encouraged to look through existing ideas to further develop or combine them, and add brand new ideas to populate the area with more sticky notes. Participants also interacted with each other through sticky notes
Sorting	5-10 minutes	Consider the factors that inform their decision making when using new technology in teaching	Participants prompted to sort sticky notes into three different areas on the Whiteboard labelled "Let's discuss this!", "Maybe discuss if we have time" and "Save for another day". Participants in both workshop sessions did not participate in this activity.
Choosing an idea from "Let's discuss this!"	5-10 minutes	Consider factors that inform decision making when using new technology in teaching	Participants asked to choose an idea out of the ones sorted into "Let's discuss this!" to create a lesson plan out of. Participants in both workshop sessions chose an idea directly from the brainstorm area.
Develop lesson plan	15-20 minutes	Reveal teachers' priorities in applying TTS to pedagogical contexts	A blank space was set up on Zoom Whiteboard for participants to take notes and create lesson plan. They were given as much time as needed to collaborate. The facilitator supported participants in using the technology when needed.
Workshop debrief	10-15 minutes	Gain extra feedback on the TTS feature in an applied context	Questions: (1) How feasible is it to run this lesson/ activity? (2) What are some possible logistical barriers you might run into? (3) What do you hope students will gain from this lesson/ activity? (4) Do you think students would enjoy the lesson?
Personal reflection	10-15 minutes	Reflect on workshop, bring specific personal and professional experiences into the conversation, additional opportunity for feedback on the TTS feature	Questions: (1) How did you find the workshop? What did you learn? (2) What are some unexpected challenges you came across during the planning process? (3) Is there anything you would do differently if you participated in this workshop again? (4) Does your perspective on the learning platform and TTS feature change when you consider different parts of your identity? (5) How can TTS be made culturally relevant to you? (6) What would you like to see us change, improve or build for the app and the TTS feature?

C Group 1 Workshop Product



D Group 2 Workshop Product