FINAL REPORT: *iPHASIA* MOBILE APPLICATION
[2483 words]

presented to
Professor Jonathan Rose

for
Creative Applications for Mobile Devices
ECE1778

by
Specialist: Stéphanie Lavigne [1003990354]
Programmer: Charvi Choksi [1006952317]
Programmer: Matan Winstok [1001130199]

University of Toronto
April 21, 2021

FINAL REPORT

1. Introduction

1.1 What

The goal of the app *iPhasia* is to help people living with Broca's (non-fluent) aphasia recover functional speech. Functional communication refers to expressing basic wants and needs in activities of daily living. *iPhasia* is a neurorehabilitation platform that aims to augment therapist-client functional outcome. It provides clients with continuous and virtual access to the speech technique Melodic Intonation Therapy (MIT) under the guidance of their therapist.

1.2 Why

Aphasia is a debilitating language impairment that often follows a stroke (Wortman-Jutt & Edwards, 2019) and 1 of 3 stroke survivors will acquire it (Aphasia Institute, 2015). Broca's aphasia is one subtype in which expressive language is impaired (Fontura et al., 2012). MIT is an effective technique to address speech fluency and research has demonstrated that MIT increased the number of words spoken per minute (Fontoura et al., 2012). MIT relies on two neural mechanisms. First, left-hand auditory-motor synchronization increases muscle coordination involved in speech production through the unaffected right brain hemisphere. Second, singing engages a different neural network than the one involved in speaking, which is likely to be intact in Broca's aphasia. Additionally, MIT uses high repetition along singing intonation (Schauld et al., 2008; Thaut et al., 2014).

2. Statement of Functionality

The main screen (**Figure 1**) provides access to the MIT "Training" platform, the metronome feature, the user's progress and the therapist input "Add a New Phrase". The "About" section provides information on the scientific evidence and the team behind the application.

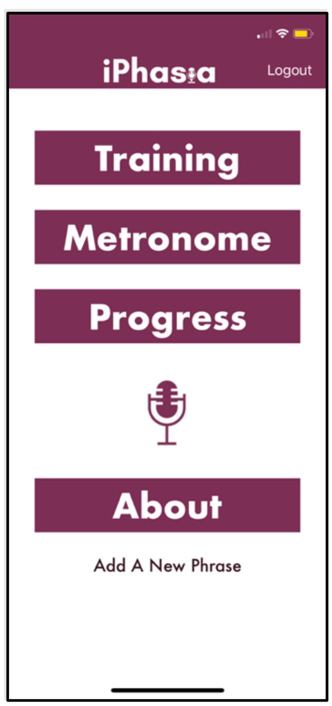

Figure 1. The main menu of the application.

The “Training” platform provides the set of training sentences specific to a user (**Figure 2**). The training material includes generic sentences provided to all users and unique sentences added by a therapist. A sentence can be selected for use in the following 6-step MIT protocol.

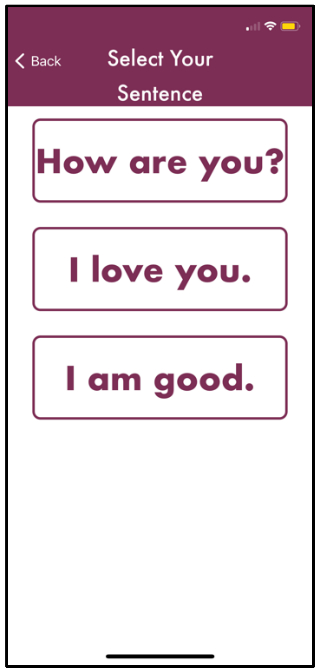

Figure 2. The set of training sentences available to a user in the training platform.

Upon sentence selection, the metronome feature provides a steady auditory beat (default is 60 BPM/1 Hz). A first pop-up instructs the user to listen to the metronome that will play for 5 seconds (**Figure 3**).

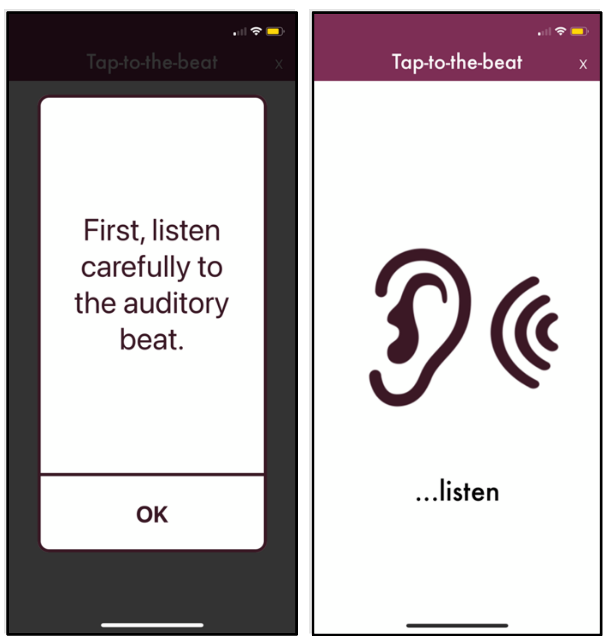

Figure 3. Left: The first pop-up instructs the user to listen to the upcoming metronome beat. **Right:** A steady beat is played and the user listens. This primer facilitates the upcoming auditory-motor synchronization task.

A second pop-up instructs the user to tap in synchrony with each beat using their left-hand index (**Figure 4**). Each tap on the screen generates a mobile device tactile haptic feedback (default is a single strong vibration). A minimum threshold of accurate taps, defined with respect to the metronome inter-beat and user inter-tap alignment periods, must be achieved to allow the user to move to the next step.

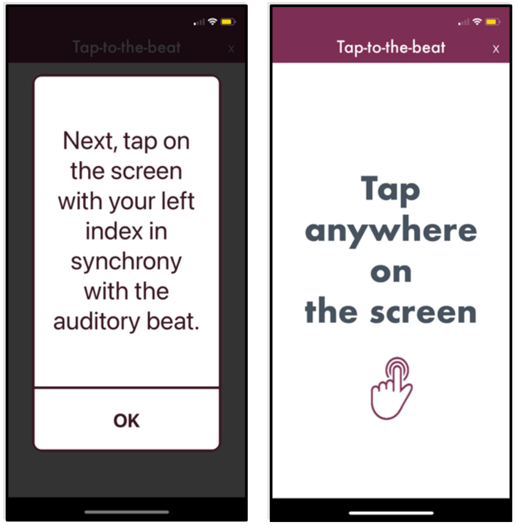

Figure 4. Left: A second pop-up provide instructions for the auditory-motor synchronization tapping task. **Right:** A steady beat plays and haptic feedback is provided through the mobile device upon each tap on the screen.

Next, a pop-up instructs the user to continuously tap along the previously heard beat tempo as it stops playing. Tactile haptic feedback remains available. Upcoming audio recordings are played along the previously tapped-along tempo (**Figure 5**). First, a hummed melody is heard by the user who listens and taps along its underlying tempo. This step is repeated 3 times. Second, a sung sentence is played and the user listens and taps along. This step is repeated 3 times. Each audio recording is spaced with a 3-second pause.

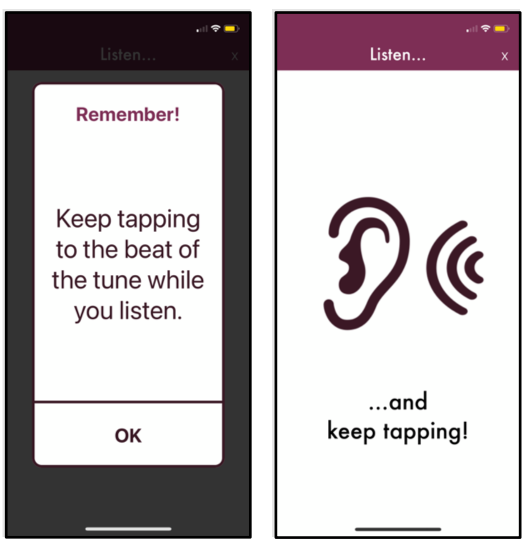

Figure 5. Left: A pop-up reminds the user to keep tapping on the screen to the underlying beat tempo with their left-hand index. The metronome stops playing but tactile feedback remains available. **Right:** The user listens to hummed and sung sentences while tapping to the underlying beat tempo.

The next step involves the user verbally repeating along an increasingly fading sung sentence (**Figure 6**). In the first repetition, the user sings along the entire sentence. The next iteration presents the sentence with a 25% fade from its end as the user independently completes it. At each iteration, the fading percentage increases by 25% of its original length. In the fourth and last presentation, 25% of the sung sentence is played from its start. This serves as a cue to assist the user in speaking the sentence independently while increasing their memory recall. Each stimuli presentation is spaced with a 3-second pause.

Figure 6. The user speaks the full sentence in synchrony to the audio recording that fades by 25% of its initial length at each repetition.

The next step of the training protocol engages the user in turn-taking. The fully sung sentence is presented while the user listens. After, the user repeats independently with no cue (**Figure 7**). This step is repeated 3 times. Each repetition is spaced with an increasing pause duration meant to trigger the working memory load. This increased information retention time intends to trigger longer term memory encoding.

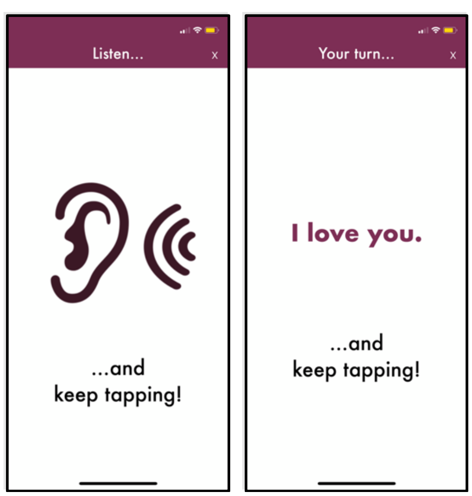

Figure 7. Left: The user listens to the fully sung sentence while waiting their turn. **Right:** The user takes turn in speaking the full sentence with no cue.

The MIT training platform ends with an assessment of the user's functional transfer (**Figure 8**). This evaluates a user's ability to employ a sentence in a real-life scenario, as this is the main goal of a neurorehabilitation intervention. A typed scenario is provided after which a simple sentence is played to prompt the user to reply with their sentence. After 5 seconds of undetected speech, the app computes an accuracy score. A pop-up provides indirect constructive feedback related to the score in the form of a string of text.

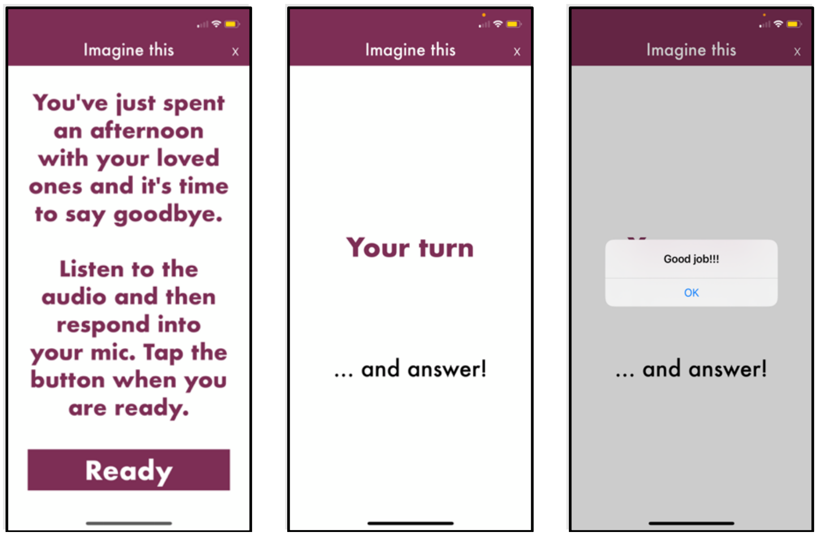

Figure 8. Left: A test scenario provides the user with a realistic everyday-life context. The "Ready" button plays a simple sentence to prompt the user to say their sentence. **Center:** The user's voice is captured and recorded through their mobile device's microphone. **Right:** The app computes a speech accuracy score and returns a string of text as indirect constructive feedback.

Upon dismissing the performance feedback pop-up, the user's performance will be updated in the "Progress" section. This section allow users to monitor their progress and needs (**Figure 9**). Tapping on a sentence will playback the last captured trial from the assessment step.

Figure 9. The progress board displays the learning progress of a user for each sentence by filling its microphone trophy with a gold colour filling (implementation intervals are: 0-24%, 25-49%, 50-74%, 75-100%).

The “Add a New Phrase” button allows a therapist to input a new sentence and a testing scenario in a user account (**Figure 10. Left**). An adjustable metronome is provided along tactile feedback upon each screen tap. This allows a therapist to practice and select the appropriate slow tempo for recording the training material (**Figure 10. Right**).

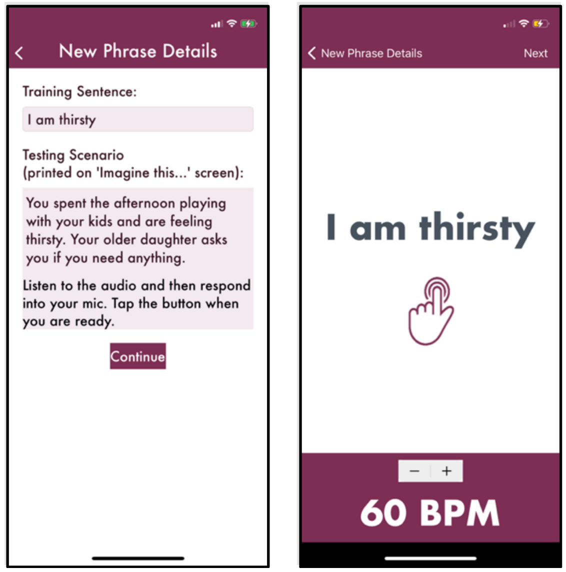

Figure 10. Left: Screen on which a therapist inputs a new sentence and a related test scenario for their client. **Right:** An adjustable metronome provides a therapist with the ability to practice and record a new sentence at a desired slow tempo.

The next screen allows the therapist to record 3 audio samples including the hummed melody, the sung sentence and the test scenario’s verbal prompt (**Figure 11. Left**). Each speaker icon to the right plays back the last recorded sample. Each sample can be re-recorded (**Figure 11. Right**). The “Save” button captures all inputs into a user account. The new sentence becomes available to this user in the sentence selection and progress board.

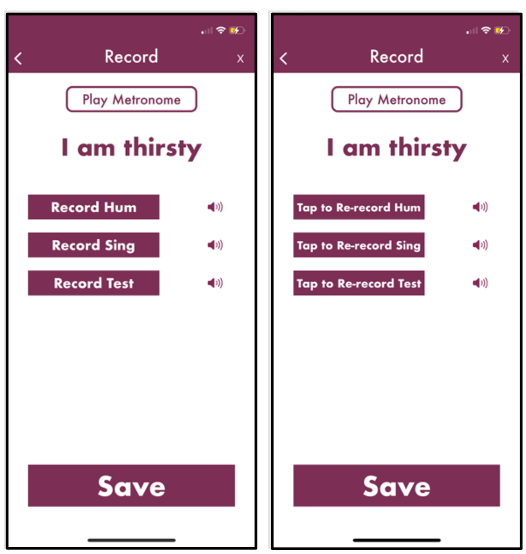

Figure 11. Left: Screen on which the therapist records 3 audio sample. “Play Metronome” provides access to the metronome. **Right:** The recorded audio samples can be played back and re-recorded, if needed.

2.1 Parts that did not work as expected
The team first explored the use of the third-party library *midi2voice* (https://github.com/mathigatti/midi2voice) to automatically generate sung melodies required in the training platform. The library takes a Musical Instrument Digital Interface (MIDI) file containing melody information and a text file as input, merges them to output a sung sentence. These audio recordings were deemed too robotic and not appropriate replicates of human performance for use in a speech neurorehabilitation context. Due to the time constraints of this project, audio samples were designed and recorded by the specialist.

3. Overall Design

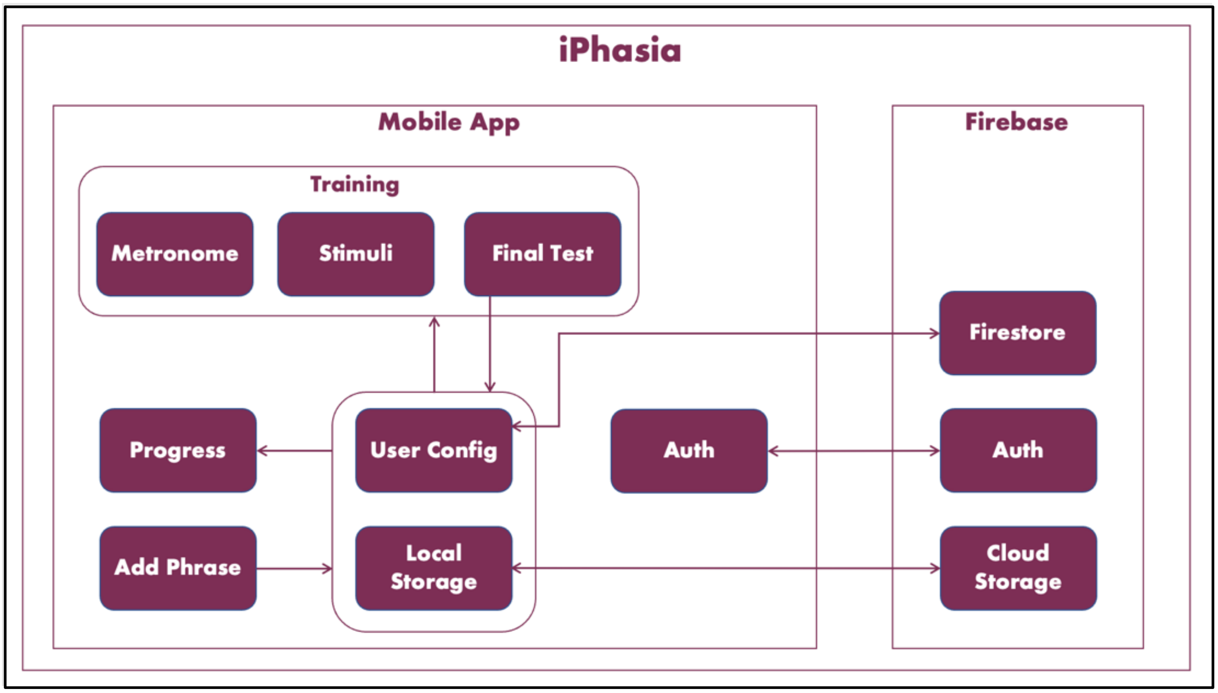

Figure 12. The "back-end" software structure of *iPhasia*.

Metronome: This feature encapsulates the rhythmic entrainment mechanism. As users tap along a steady auditory beat, their motor system couples with their auditory system. In the stand-alone metronome feature, users can adjust the beat tempo. In the training platform, the metronome is set to the tempo of the sentence. In the therapist feature, the metronome tempo is adjustable. Tapping accuracy consists in the user inter-tap period and metronome inter-beat period aligning within a <10% deviation window, 80% of the time.

Training: This feature includes selecting the sentence to be used in the 6-step MIT protocol. The integrated metronome provides auditory-motor entrainment. This feature serves as a training platform to provide auditory playback, tactile feedback and captures a user's voice during the test scenario. The captured voice is analysed for accuracy, reflecting speech intelligibility.

Progress: This single screen displays a user's performance on each phrase. Performance for a sentence is represented by the amount of gold colour filling its microphone trophy displays. Filling is implemented by intervals of 25% according to the test score for speech intelligibility. The user's last captured audio test can be played when tapping on a sentence.

Add Phrase: This feature allows a therapist to input a personalized sentence into a user account. A sentence and realistic scenario is typed in and a tempo is selected with the integrated metronome. This allows a therapist to select the appropriate tempo for recording the hummed melody, sung phrase and spoken testing prompt through the mobile device's microphone. The new sentence is added to the local and cloud datastore and is made available to the user in their training sentence set.

Datastore: User data, including sentences and progress, are stored in an internal module called UserConfig. Audio files are downloaded and stored in the phone's local storage.

Auth: Authentication allows the user to access their data from multiple devices.

Firebase: Firebase Auth, Firestore, and Cloud Storage are used for performing user authentication and storing user data and audio files.

4. Reflection

The team performed a retrospective analysis on what went well, what didn't go well and what could have gone better.

4.1 What Went Well

The team was cohesive and efficient in task distribution. Communicating in an inter-disciplinary space was an ongoing learning process which proved to be fructuous.

4.2 What Didn't Go Well

The algorithmically generated singing voice did not meet the project requirements which forced the team to redesign the implementation of this feature.

4.3 What Could Have Gone Better

The therapist as an app user was incorporated mid-way into the development of the project. The therapist feature and its extensions are areas that will require more work to be fully developed. The main challenge faced by the team was in translating synchronous face-to-face multi-sensory interaction into a human-to-mobile device interaction. This includes synchronous auditory, visual, tactile and proprioceptive sensory stimuli.

5. Group Contributions

Matan (programmer) implemented the MIT steps involving audio playback, recording, and repetition. He worked on the therapist feature ("Add A New Phrase") and integrated Firebase into the app. He performed research and development (R&D) to implement string similarity algorithms for generating a speech accuracy test score.

Charvi (programmer) implemented the registration and login, the metronome and the progress features, as well as parts of the training flow. She worked on the database design and the base user interface (UI), navigation, and data flow. She performed R&D on the computer-generated singing voice, although this option was not carried forward.

Stéphanie (specialist) provided the conceptual design of the app to serve as a neurorehabilitation tool. She provided background information on aphasia, the MIT protocol and its underlying mechanisms. She guided the sensory feedforward-feedback loop implementation. She developed the UI app prototype with considerations to older adults as main target users. She designed the logo and trophies. She recorded scenarios, sentences and melodies, and assisted in the testing and improvement of the app.

6. Specialist Context

MIT was developed by neurologic researchers in the early 1970s (Schauld et al., 2008; Thaut et al., 2014) but until today, its full software implementation has not come to our attention. Due to the COVID-19 pandemic, Neurologic Music Therapists (NMTs) who transitioned their therapy services from in-person to telehealth were successful in addressing all NMT techniques, with the exception of rhythmic auditory stimulation. Likewise, the inability to physically guide hand-tapping in the MIT protocol was noted as a limitation (Cole et al., 2021). Therefore, this app is suited to address the need for synchronous auditory, visual and tactile stimuli, although more R&D will be required to deliver remote proprioceptive cues. Additionally, the app could augment the therapist-client functional outcome by providing a client with a continuous and virtual access to their personalized MIT protocol. A client could independently, or with assistance of their caregivers, practice between therapy meetings. For research, this app could serve as a clinical tool, i.e., to deliver and monitor a participant’s accuracy during different auditory-motor synchronization finger-tapping tasks. Additionally, these tasks could be conducted remotely.

Virtual access to this neurorehabilitation tool could help the provision of therapy services to individuals living in remote areas and to those with fragile health. It could help demarginalize access to therapy by providing a lower cost virtual access to a neurorehabilitation tool.

7. Future Work

The team intends to develop this app beyond the context of this course. *iPhasia* will be presented to NMTs, Speech and Language Pathologists (SLPs), rehabilitation specialists and other health professionals through different communication channels, including accreditation boards and associations.

The team’s upcoming efforts will gravitate around augmenting the therapist feature. First, the therapist “Add A New Phrase” option, currently actionable in-person through a client’s profile, will be extended to a remote option. Second, as some SLPs may not be musically trained, we will implement an algorithmic melody generator that will match each sentence syllable length and type in order to replicate natural prosody. Third, we intend to implement playable musical notes on the therapist UI, i.e., as a virtual music keyboard, to help musically trained therapists rehearse and generate their own melodies. Fourth, the training UI will display visual facial and lip movements involved when speaking a sentence. The therapist will be offered the option to record themselves speaking a sentence using the front camera of their mobile device. This visual component captured and displayed in-sync with the auditory file will provide clients with an additional means of learning through imitation, thus increasing the potential outcome of the treatment.

8. Agreement to Post Publicly

All group members have agreed to publicly share the report and final presentation video on the course website. We ask that the source code does **not** get publicly posted.

References

American Speech-Language-Hearing Association. (n.d.). *Aphasia.* Retrieved February 12, 2021, from https://www.asha.org/practice-portal/clinical-topics/aphasia/.

Aphasia Institute. (2015). *What is Aphasia?.* Retrieved February 12, 2021, from https://www.aphasia.ca/home-page/about-aphasia/what-is-aphasia/.

Cole, L. P., Henechowicz, T. L., Kang, K., Pranjic, M., Richard, N. M., Tian, G. L. J & Thaut, C. H. (2021). Neurologic music therapy via telehealth: a survey of clinician experiences, trends, and recommendations during the COVID-19 pandemic. *Frontiers of Neuroscience, 15*. doi: 10.3389/fnins.2021.648489.

Fontura, D. R., Carvalho Rodrigues, J., Sa Carneiro, L. B., Monçao, A. M. & Salles, D. F. (2012). Rehabilitation of language in expressive aphasias. *Dementia & Neuropsychologia, 6*(4). pp. 223-235. doi: 10.1590/S1980-57642012DN06040006.

Li, J. (2016). *StringMetric* [Source code]. Retrieved from https://github.com/autozimu/StringMetric.swift/.

National Aphasia Association. (n.d.). *Aphasia Definitions*. Retrieved February 12, 2021, from https://www.aphasia.org/aphasia-definitions/.

Racette, A., Bard, C. & Peretz, I. (2006). Making non-fluent aphasics speak: sing along! *Brain, 129*. pp. 2571-2584. doi: 10.1093/brain/awl250.

Schlaug, G., Marchina, S. & Norton, A. (2008). From singing to speaking: why singing may lead to recovery of expressive language function in patients with broca's aphasia. *Music Perception: An Interdisciplinary Journal, 25*(4). pp. 315-323. doi: 10.1525/MP.2008.25.4.315.

Stahl, B., Kotz, S. A., Henseler, I., Turner, R. & Geyer, S. (2011). Rhythm in disguise: why singing may not hold in key to recovery in aphasia. *Brain, 134*. pp. 3083-3093. doi: doi:10.1093/brain/awr240.

Sun, X. (2020). *XSMetronome* (Version 1.0) [Source code]. Retrieved from https://github.com/xiangyu-sun/XSMetronome.

Thaut, M. H., Thaut, C. P. & McIntosh. (2014). Chapter 11: Melodic Intonation Therapy (MIT). In Thaut, M. H. & Hoemberg, V. (Eds.), *Handbook of neurologic music therapy* (pp. 140-145). Oxford University Press.

Wortman-Jutt, S. & Edwards, D. (2019). Poststroke aphasia rehabilitation: why all talk and no action? *Neurorehabilitation and Neural Repair, 33*(4). pp. 235-244. doi: 10.1177/1545968319834901.

Zumbansen, A., Peretz, I. & Hébert, S. (2014). The combination of rhythm and pitch can account for the beneficial effect of melodic intonation therapy on connected speech improvement in Broca's aphasia. *Frontiers in Human Neuroscience*, *8*(592). pp. 1-11. doi: 10.3389/fnhum.2014.00592.

Contacts

Specialist
Stéphanie Lavigne, PhD Student, MA, MT-BC, MTA, NMT
Music and Health Science Research Collaboratory, University of Toronto
stephanie.lavigne@mail.utoronto.ca

Programmers
Charvi Choksi, MEng Student in Computer Engineering
University of Toronto
charvi.choksi@mail.utoronto.ca

Matan Winstok, MEng Student in Computer Engineering (Analytics)
University of Toronto
matan.winstok@mail.utoronto.ca