UX Research Report for Pera

Krishna Cheemalapati, Tony Juntao Hu, Leonid Nediak, Wilson Sy, Parth Vats

1 Method

1.1 Participants

We recruited one U of T student from Robarts Library, who we identified as a potential user of our application based on our initial screening questions, discussed in section Section 1.3. We had originally planned to recruit up to 4 students. However, after the first interview, we identified a few significant problems with our materials which must be fixed for any subsequent interviews to be effective. As a result, and due to time constraints for this assignment, we postponed further interviews.

Additionally, we conducted the same interview process with the course instructor. The materials presented were improved based on what we discovered during the student interview.

1.2 Materials

We built a low-fidelity "paper prototype" using Google Slides. The source slides can be found here. The prototype implements three use cases, namely "Reading Karaoke", "Long Reading", and the progress dashboard. For the UX research, we printed the slides and used them as if they were a real paper prototype drawn by hand. However, for simplicity, instead of re-using components such as sidebars, we simply printed each screen containing those components. The only exception was the tooltip pop-ups, which we printed and cut into individual pieces. When interviewing the instructor online, we used Miro as a virtual board and displayed images exported from the slides in place of paper printouts.

1.3 Procedure

We first approached potential participants by using the following introduction: "Hi there, we're building an application for second language learning and we're currently doing some research for it. We are just wondering…" We then asked the following screening questions

1. Are you currently learning a language?
2. Are you struggling with pronunciation in this language?
3. Are you free for 10-15 minutes to try out our prototype?

If the participant answered "yes" to all of the above, then we invited them to sign the usability consent form, found here.

Two experimenters were present in each interview session. One experimenter read aloud the context script, which contains some background information and instructions (see Appendix A). The same experimenter then presented three scenarios for the participant, one at a time (see Appendix B). After being presented each scenario, the participant interacted with our prototype while being video-recorded. The second experimenter acted as the computer and responded to participant input by swapping in and out the relevant screens. No additional help was provided except when the participant was confused about the task itself.

After completing all three scenarios, participants filled out a System Usability Scale (SUS) questionnaire, found here. Finally, we conducted a post-task interview with a few guiding questions to collect additional feedback (see Appendix C).

2   Results

Usability testing for Pera revealed significant issues in navigation, feedback comprehension, and UI design. Users struggled to differentiate between Reading Karaoke and Long Reading modes and found the feedback (IPA symbols in particular) confusing and unhelpful. The Long Reading feedback screen left users unsure of how to interpret results, and was very text-heavy. The process for adding passages to Reading Karaoke and generating new ones was unintuitive, causing frustration. UI elements like the ear button were overlooked, and dropdown menus and progress indicators did not function as expected.

Key improvements are needed. A tutorial or onboarding guide should clarify app functionality for new users. Critical UI elements, such as the record button, need better visibility, and should better indicate when the application is recording. The feedback system should replace IPA with clearer explanations and highlight mistakes within a general window of where the error occurred within the user's response. The generate screen should offer random/popular prompts with an animation cycle for easier selection. Language selection should move to settings, and button placements need optimization to improve accessibility.

Compared to competitors, Pera was seen as a serious pronunciation tool for intermediate and advanced learners. To remain competitive, however, Pera must refine its UI, clarify feedback, and ensure a smoother user experience, especially for new users.

Raw notes from the sessions are included in Appendix D.

Appendix A   Context Script

Welcome to our usability study, and thank you for your time. Our project is designed to help individuals learn languages with highly personalized and AI augmented feedback. Our system is specifically targeting people who are working on building their pronunciation abilities, as there is a lack of strong existing solutions for such a product at this time.

We have recruited you for the purposes of this design evaluation because you indicated that you are learning a new language and indicated that you have struggled with pronunciation. We assume that you are comfortable using mobile and desktop apps and voice recording technology.

In this study, we will give you three tasks and ask you to navigate the app as you see fit to complete each task. During the tasks, my colleague is going to act as the computer and manipulate the screens based on your inputs. If he does not respond, assume that the functionality is unimplemented. Due to the nature of these tasks, we will not provide additional instructions or answer questions related to the application.

Finally, please verbalize your thought process along the way.

Appendix B Scenarios

Scenario 1A: You want to practice your pronunciation in a language you are trying to learn, but don't have a specific passage to read. You've just signed up for a new language learning tool and have been sent to the dashboard. **Show me how you would generate a passage using AI and start practicing the pronunciation of individual sentences in that passage with the Reading Karaoke feature.**

Scenario 1B: You want to practice your pronunciation in a language you are trying to learn, and you have a specific passage you want to practice reading aloud, such as a news article or a book excerpt. You've just signed up for a new language learning tool and have been sent to the dashboard. **Show me how you would import your own text and start practicing the pronunciation of individual sentences in that passage with the Reading Karaoke feature.**

Scenario 2A: You want to improve your fluency by practicing reading longer texts, such as an article or a short story. You have a document you'd like to use for this purpose. **Show me how you would import a long text and begin practicing with Pera's Long Reading mode.**

Scenario 2B: You want to improve your fluency by practicing reading longer texts, such as an article or a short story, but don't have a specific text in mind. Pera allows you to generate a passage tailored to your proficiency level. **Show me how you would use AI to generate a long text and begin practicing with Pera's Long Reading mode.**

Scenario 3A: You previously practiced a passage using Reading Karaoke and want to revisit it to track your progress. **Show me how you would find and resume a past Reading Karaoke session.**

Scenario 3B: You previously practiced a passage using Long Reading and want to revisit it to track your progress. **Show me how you would find and resume a past Long Reading session.**

Appendix C   Interview Questions

1. Were there any frustrations or areas of confusion while navigating the tasks?
2. How did you perceive the utility of the feedback you are receiving when completing the tasks?
3. Would you use this solution over existing solutions such as Duolingo and why?
4. Are the options provided in the settings and dashboard useful? Is there anything missing?
5. Do you have any suggestions for improvement to the app or overall experience?

Appendix D Raw Notes

Participant 1

Learning Mandarin and Russian.

Notes during task:

- chose long reading instead of reading karaoke when prompted with scenario 1a
- spend a while deciding in the generate screen
- we didn't have the pre-record screen for long reading?
- not sure how to find feedback after finishing recording, pressed help button
- confused about what "repetition" means in long reading feedback, and about the feedback in general
- repetition feedback tooltip not clear
- if the button name is "generate new passage", it should go to the generate functionality instead of the welcome screen
- believes she has to generate passage first (through long reading welcome screen before going to reading karaoke
- feel like when adding to reading karaoke through long reading, should not be brought directly to reading karaoke functionality, but should just be told "added"
- She did not see the "ear" button
- feels like IPA doesn't do that much (Tony: I was going to add sounds and illustrations to each of those IPA symbols)
- feel like teleprompter option should be greyed out if not turned on?

Were there any frustrations or areas of confusion while navigating the tasks?

> There was a time when i added the long passage to the karaoke i was kinda surprised / confused why it got moved to the karaoke screen right away

> idk if its a confusion but i found it tedious to press generate new response and get sent to the thing and then have to click it again in the middle

> Don't use IPA in feedback for anything

> Confused as to what reading karaoke and long reading were, later tried to find out what each mode was by going to the help page but it didn't have anything, not that i think it has to be there but I couldn't really find a place describing what each thing was

So like a tutorial?

> Yeah like a stuck or a how to use?

Was the question mark not sufficient? Should we have multiple help menus?

> Have short description of what the mode was somewhere, the placement of the help text is a design choice, maybe if new they'd want to see the help right away but if they're familiar then they won't wat to see it all the time.

> The screen feels like it's assuming I'm a returning user.

> Maybe recording should say "press again for feedback". For some people it might be taken for granted computer will automatically and "intelligently" stop recording and generate feedback

When pressed the dropdown button for progress, expected a dropdown menu, but was surprised that she was brought to a new page. Was expecting to see sentences she said with "..." if cut off.

Expected sentence by sentence view in dropdown, and not necessarily the whole paragraph.

How did you perceive the utility of the feedback you are receiving when completing the tasks?

Pretty useless, because IPA.

Although, if you provide a sample word as reference, maybe they say the sample word wrong too.

The long reading feedback screen confused her a lot. Wasn't sure how "rephrasing" can work if she's just reading the text (explanation from the tooltip).

The ability to re-listen to your recording might be helpful. Maybe have a bar for replaying the recording, and have the bar coloured to show where the mistake happened.

Give a "window" for where the mistake happened rather than the exact location.

It'll be hard to colour a mistake if multiple mistakes happen at the same location.

Have multiple parallel lines, each for a different kind of mistake.

Feels like people will not believe that they said filler words if they see the "filler word" feedback.

Importing, she had to say the type of file that needed to be imported. She thought it was for file type selection. Maybe put text input region and drag-and-drop functionality on the same screen. Dislikes the dropdown menu.

Did not realize the ear was a clickable

Would vowel visualization be useful?

Not really, but maybe have an "advanced mode" that users can opt into. This needs to be taught. Maybe it can be part of a "premium version".

Would you use this solution over existing solutions such as Duolingo and why?

Doesn't have specific feedback and just says "you got a word wrong". A lot of the time she disagrees with the feedback.

Loves our idea of longer paragraph vs. shorter sentence mode. In the long ones you get cadence, intonation, etc. In the short one it's words, pronunciation, am I getting things right?

Would need to be a mobile app for her to choose it over other solutions.

Are the options provided in the settings and dashboard useful? Is there anything missing?

Doesn't know what counts as a major mistake.

Suggestion: Include checkboxes for different types of errors, hide

Do we need a save button for options / settings?

Do you have any suggestions for improvement to the app or overall experience?

don't need to include default subject for generate

include random / popular in generate section, user voting for effectiveness of generated passage?

Use an animation to show random prompts in the generate text box, change every 5 seconds.

Teleprompter feels like karaoke?

Having a fun name (karaoke) is important, as it makes people have fun using the application. But fun name may not be accessible or comprehensible to everyone.

The learning language and base language should be in settings, instead of in the sidebar. Look at Duolingo and italki for how they expose language options.

Course Instructor

Notes during task:

- mistook drag and drop window for text input box
- didn't press record when first going into the reading screen for reading karaoke
- didn't stop recording when she stopped speaking in reading karaoke
- did not discover the progress overview page, but knew that she could change the sentence numbering by typing
- assumed for a split second tat feedback strictness and the other slider were click-based, but realized they needed to be dragged.
- did not click record for long reading, and did not click end record?

Were there any frustrations or areas of confusion while navigating the tasks?

forgetting to press record, maybe if it's blinking or something then it's better

setting up the import and configuring the difficulty, difficult to see the contrast of the sliding bit (dial/handle) and the other bits

start button is maybe… a little far away?

How did you perceive the utility of the feedback you are receiving when completing the tasks?

liked the reading karaoke, liked that she could click on the "play" buttons

long reading definitely was a bit more overwhelming

not sure what the colours mean but figured it out in the interview stage (now)

Very text heavy interface for long reading (initial impression), but she says she understands that's the nature of this task. Maybe can see a high level overview first, but maybe that's too much clicking

maybe too many options presented at the end of long reading (3 buttons), maybe if there's one call to action that would be better

Would you use this solution over existing solutions such as Duolingo and why?

For pronunciation, yes, because Duolingo is not pronunciation focused, but pre-requesite, if I'm intermediate proficiency already.

She perceived our product as more serious, so says it's for ppl who are "advanced", important to feel like you're improving and having a feedback loop

Are the options provided in the settings and dashboard useful? Is there anything missing?

Would have thought settings would be the place for history, had she not done the long reading task first.

Thinking out loud, does changing the default settings affect the dashboard (e.g. leaderboard) and stuff or does it not impact existing settings?

On the dashboard, overall score is a bit confusing. However, would want to see dashboard like this.

Do you have any suggestions for improvement to the app or overall experience?

N/A

Did the whole progress bar stuff make sense?

The surprise feature (progress overview in the reading karaoke) was very hidden, but base progress was clear.

Understood that "grey" means currently on, but didn't immediately perceive the orange as meaning anything, just thought it was filled in progress.

Likes that having "red" on a sentence doesn't block you from advancing, but still likes that it shows red.

Would the colours (on the progress overview page) be excessive?

Colours would take a few iterations to get right.

Questionable that whole sentence is highlighted when only individual words are wrong. Maybe only highlight the words that need feedback, or another way (I didn't catch what Michelle said).