Final PRD
Generative AI ESTM 60244 | Sarah Noel

Product Overview

In the United States, 180,000 individuals are diagnosed with aphasia each year and 2 million Americans currently live with communication deficits. The most common causes of aphasia are strokes and traumatic brain injuries (TBI), however, any condition that damages your brain can impact the language center and diminish your ability to speak and communicate. While patients receive intensive therapy during hospitalization, access drops dramatically once they return home due to insurance limits, high cost, limited availability of speech-language pathologists, and transportation barriers. Without continued practice, progress plateaus, motivation declines, and long-term outcomes suffer. Caregivers assume the burden of home care without tools or expertise, and patients often lose confidence in their ability to communicate.

This platform addresses this gap through an AI-powered speech and language recovery assistant designed for individuals with mild to moderate aphasia or speech impairment. It allows users to engage in structured speech practice, receive AI-generated feedback, and simulate conversational exercises that enable more frequent, flexible rehabilitation outside the clinic setting. The platform adapts complexity as a user improves, supports short daily sessions for cognitive fatigue, and provides mock progress summaries to reinforce motivation.

Core Features

Feature	Status	AI Use
Accessibility Features	Implemented	No
Guided Exercises	Implemented (prototype)	Yes
Difficulty Progression	Implemented	Yes
Conversational Simulation	Partial	Yes
Report Export	Partial	No
Pronunciation Scoring	Implemented	Yes

AI Specification

Task	Description	Input	Output	Model
Speech Analysis	Recognition and evaluation of user's voice input. Metric and pronunciation feedback outputs	Audio file	Accuracy %, pacing notes	Gemini Multimodal API

	allow for corrective feedback and changes in-real time			
Adaptive Drills	Recommendation and generations made based on histories of performance and creates an output of the modified difficulty in the following exercise. This allows for therapy to be kept at a challenging but attainable level.	Error rate, response quality	Easier or harder follow-up	Gemini API
Conservation practice	Generation of unique and non-repeated topics or scenarios that follow natural dynamic dialogue to build confidence in real work applications.	User speech response	Next prompt, correction, guidance	Gemini API
Session Summary	Produces recap + next steps	Session text log	Progress summary	Gemini API - Summary Generation

User receives prompt → User responds → AI evaluates → AI gives feedback → repeat → summary generated.

Guardrails: No medical diagnosis provided through the platform, storing of speech and text data for reference in summary must be have consent
Guardrails to be added later: Content filters for unsafe/off-topic language

Technical setup

Google AI Studio will be used to create a prototype of the platform. Conversational and feedback logic, generative adaptive dialogue and processing of user input audios will be tested here through prompt engineering. The frontend user interface will include a ‘start/stop session’ interaction. Audio recording will capture the user's input and send it to be processed. Metrics created to show session progression/duration and scores will be displayed in the environment. Google Multimodal will take the real time audio recording, recognize and create error correction to guide adaptive prompts. Prompts should mimic a SLP. Report generation will be done by a generated dashboard that summarizes session progress and trends overall. Ability to export summaries must be possible. GitHub Pages allows for live deployment and public access. There is no long-term data storage. Data is handled per session.

Prompting strategy

“Generative a responsive HTML landing page for a healthcare platform to aid in speech language recovery. Include description of the service, a menu to view ‘past session reports, export report, goal progress, set a new goal’ and button to start a new session”

“Simulate an interactive speech therapy session when the AI acts as a speech language pathologist. Provide a short, clear prompt with a question. Offer gentle feedback in pronunciation or cadence is not accurate.”

“Ajust following prompts’ difficulty based on the performance score of the response just given.”

“Summarize this session into a user report that highlights areas that have seen progress and a couple realistic potential next goals to set to continue improvements.”

“Insert placeholders where real pronunciation-analysis API will be integrated. Add comments labeled // future speech-to-text scoring here in code.”

“Add color-coded phoneme scoring UI:
Green = correct, Yellow = approximate, Red = distorted or missing.
Output results in a simple bar or grid visualization.”

“WAIT for an audio file and target phrase before responding.”

“Target phrase: "The red apple is on the table."
Audio input: (user speech recorded)

Expected Model Output Example:
"transcription": "The red apple is on table",
"accuracy_score": 82,
"word_match_percent": 90,
"fluency_notes": "Slight hesitation before 'table'.",

"pronunciation_feedback": "Great effort! Try exaggerating the ending sound in 'apple' and adding 'the' before 'table'.",
"next_practice_recommendation": "Repeat phrases again once, then try increasing speed slightly."
”

Overtime prompts became more focused on one change at time. I started providing examples of outputs I wanted. I learned to move away from vague responses (shorter or easier to specifically say 1-2 lines of feedback max). Stating from the beginning that this was not a platform to replace SPLs, and should never make medical claims or use diagnostic language was important, but in later prompts I made sure it was stated again.

UX & Limitations

Intended user Journey:

- User opens the platform and selected either Guided Drills or conversational practice mode
- AI presented a prompt
- User responds through speech input
- AI generates feedback
- The next prompt is determined based on performance. Repeat 2-5.
- The session concludes with a summary.
- Progress is tracked and carried over from sessions

Known Limitations:
User starts fresh each visit because progress does not persist between sessions
There are no constraints on conversational mode

Users should not rely on the platform for medical diagnostic information, determining extent of aphasia or treatment needs, patients requiring clinical assessment must seek an SLP. The goal is to be therapy adjacent not replacing.

Future Roadmap

- Creating an opt-in storage of session data. Ideally data would be encrypted. This would also have a layer to automatically input records into EMR to be referred to.
- Multilingual support
- Gamifying metrics to include badges or challenges
- Clinical portal with telehealth abilities. Addition of a ‘schedule an appointment with your SPL’ that would use schedule optimization to book appointments with medical providers if needed.