Pronunciation Coach – First Milestone

Jorge L. Luna Pagán; Team 2

19 September 2025

Table of Contents

1. Informative Part

1.1. Teams

1.1.1. Team Members

Name	Role	Tasks
Jorge L. Luna	Team leader	Project coordination, research analysis of phonemes, implement Vosk as speech-to-text for testing
Daniel Reyes	Developer	Research speech-to-text models, provide findings to team, implement Whisper (OpenAI) as speech-to-text for testing
Uriel Rosado	Developer	Research pronunciation comparison methods, evaluate Montreal Forced Alignment models
Claudia Guzmán	Developer	Explore AI-based user feedback (chatbot), study phonetic dictionaries
Diego Rios	Developer	Investigate playback-based pronunciation training, creation of phonetic dictionaries
Omar Cordero	Developer	Analyze speech-to-text libraries, develop algorithms for phoneme-level pronunciation scoring
José Valentín	Developer	Evaluate STT options, research English dictionaries
Noel Colón	Developer	Research accent variation in pronunciation, build orthographic dictionary

1.2. Current Situation, Needs, Ideas

1.2.1. Current Situation

Language learning applications are widely used by students, professionals, and independent learners, but most of them focus on vocabulary acquisition, grammar, or listening comprehension rather than accurate pronunciation. Existing tools such as Duolingo or Babbel provide limited pronunciation feedback, often only a "correct/incorrect" judgment without detailed guidance.

Speech-to-text (STT) technologies have advanced significantly in recent years. Models like Whisper (OpenAI) achieve high transcription accuracy, but they are computationally heavy, while alternatives like Vosk run more efficiently on consumer devices with lower accuracy. Despite this progress, most STT solutions are not designed specifically for pronunciation evaluation, and they rarely offer phoneme-level analysis.

Learners, especially non-native speakers, face difficulties practicing pronunciation independently.

Educational institutions provide some support, but access to teachers and speech experts is limited, and many learners want an affordable, offline solution they can use on their own devices.

In summary, while robust STT technology exists, there is a gap in user-friendly, accessible, and affordable pronunciation coaching tools that provide actionable feedback. This is the environment in which the Pronunciation Coach project emerges.

For example, these are great software to learn a languege, but lack of that pronunciation feedback that the user needs: [Duolingo]: https://www.duolingo.com, [Babbel]: https://www.babbel.com

1.2.2. Needs

From the current situation, several categories of user and domain needs can be identified:

Pedagogical Needs

- Clear and actionable feedback on pronunciation beyond a binary “correct/incorrect.”
- Independent practice opportunities without constant teacher supervision.
- Adaptability to different accents and common pronunciation challenges.
- Motivation and progress tracking to support long-term learning.

Accessibility Needs

- Affordable solutions that reduce or eliminate subscription costs.
- Offline availability, allowing use without continuous internet access.
- Simple, intuitive tools that can be used by learners of different ages and technical skills.

User Needs

- Tools for learners to monitor their own progress over time.
- Mechanisms to highlight frequent pronunciation errors for focused practice.
- Transparent and explainable evaluation methods, showing why certain pronunciations are marked as incorrect or needing improvement.

1.2.3. Ideas

To address the identified needs, the following feature-oriented ideas have been considered:

Pedagogical Features

- Provide detailed feedback that highlights mispronounced words or phonemes, with suggestions for improvement.
- Offer sentence and word practice modes to let learners focus on specific areas.
- Support accent-aware evaluation so that learners with different linguistic backgrounds receive fair and useful feedback.
- Include progress tracking dashboards that visualize learner improvement over time.

Accessibility Features

- Maintain a low-cost model by relying on open-source speech recognition and feedback methods.
- Design a simple, user-friendly interface suitable for both beginners and advanced learners.

User Features

- Allow learners to monitor their own progress and identify areas for focused practice.
- Provide clear visualizations of frequent pronunciation errors.
- Ensure transparency in feedback by showing how evaluations are derived (e.g., highlighting words or phonemes instead of giving only scores).
- Enable exporting progress reports for personal review or sharing with tutors/mentors.

1.3. Scope, Span, and Synopsis

1.3.1. Scope and Span

Scope

A digital language learning app, specifically tools designed to help learners improve spoken English. This includes general language apps, pronunciation tools, and speech analysis technologies.

Span The Pronunciation Coach focuses on a specific segment of this domain: an application that provides learners with detailed, actionable feedback on their pronunciation at both word and phoneme levels. The project emphasizes accessibility, low-cost solutions, and visual progress tracking for independent learners.

1.3.2. Synopsis

Synopsis The Pronunciation Coach is a software tool aimed at helping language learners improve their English pronunciation. By leveraging open-source speech-to-text models, the application evaluates user speech at the word and phoneme levels, highlights errors, and provides clear, actionable feedback. The tool is designed to track progress over time to motivate continued practice. This solution addresses the gap in current language learning tools that often provide minimal or non-specific pronunciation feedback.

1.4. Other Activities (Beyond Coding)

In addition to core development, the Pronunciation Coach project involves several supporting activities:

Domain Engineering

- Studying language learning techniques, phonetics, and pronunciation challenges.
- Reviewing existing STT models (Whisper, Vosk) and their suitability for offline evaluation.

- Exploring **Montreal Forced Alignment (MFA)** for phoneme-level alignment and error detection, evaluating its potential for accurate feedback in pronunciation learning.

Requirements Analysis

- Identifying user needs (learners) and mapping them to feature ideas.
- Defining system requirements for accuracy, offline performance, and usability.
- Exploring algorithms that will identify the user's erros.

Architecture

- Designing the software architecture to integrate recording, STT processing, feedback generation, and progress tracking.
- Planning for modularity to allow swapping or updating speech recognition models.

Testing

- Conducting usability tests with learners to evaluate comprehension and effectiveness.
- Comparing STT outputs with target phrases to validate accuracy.
- Evaluating performance across different accents and age groups.

Deployment

- The Pronunciation Coach application should be **lightweight**, running smoothly on typical consumer devices such as laptops, tablets, and smartphones without excessive CPU or memory usage.
- The user interface should be **intuitive and user-friendly**, allowing learners of varying ages and technical proficiency to navigate recording, transcription, and playback easily.
- The system should support **offline operation** for core functions (recording, transcription, playback) to ensure accessibility in environments with limited connectivity.
- Packaging and installation should be simple, requiring minimal setup for learners to start practicing immediately.

1.5. Derived Goals

Beyond the primary objective of helping learners improve pronunciation, the project aims to achieve:

- Explore how open-source STT models can be adapted for educational purposes.
- Provide insights into pronunciation errors across different accents and linguistic backgrounds.
- Develop a framework that can be extended to support additional languages or advanced phonetic feedback in the future.
- Promote learner independence by offering a tool that works without requiring continuous teacher intervention.

2. Descriptive Part

2.1. Domain Description

2.1.1. Domain Rough Sketch

The domain of pronunciation coaching was explored through brainstorming, observations of language learners, and analysis of existing tools. Key raw notes and observations include:

- Learners often struggle with specific sounds in English, such as “th,” “r/l,” and vowel contrasts, depending on their native language.
- Many learners want immediate, actionable feedback without waiting for a teacher.
- Current language learning apps (e.g., Duolingo, Babbel) offer limited pronunciation guidance—mostly binary correctness or repetition tasks.
- Learners benefit from seeing visual representations of their pronunciation, such as waveform, pitch, or phoneme highlights.
- Speech-to-text engines like Whisper (OpenAI) provide accurate transcription but require more resources, while Vosk runs efficiently offline with lower accuracy.
- Feedback should be understandable, not just a numeric score, to help learners correct mistakes.
- Learners’ accents vary widely, requiring evaluation systems that can adapt or be tolerant to variation.
- Phonetic dictionaries and mapping of phonemes are needed for accurate feedback and scoring.
- Teachers or advanced learners may want to export or track progress for study or coaching purposes.
- Early prototypes could integrate simple dashboards showing practice frequency, error frequency, and improvement over time.
- User experience is important: intuitive interface, easy recording, playback, and comparison of speech with target pronunciation.
- Potential additional features: repetition suggestions, highlighting difficult words, or guiding learners through tongue position/phonetic tips.
- Integration with chatbots or AI feedback systems could provide more interactive, personalized learning.
- Using tools online can be hard for your pronunciation development, we need something intuitive for the user.

2.1.2. Terminology

- **Learner** – A person practicing pronunciation to improve their spoken English.
- **Pronunciation Feedback** – Information provided to the learner about the correctness or quality of their spoken words or phonemes.
- **Phoneme** – The smallest distinct unit of sound in a language; used to identify specific

pronunciation errors.

- **Word-Level Accuracy** – Measure of correctness for individual words in a sentence.
- **Speech-to-Text (STT) Engine** – Software that converts spoken audio into written text, e.g., Whisper (OpenAI) or Vosk.
- **Offline Mode** – Ability of the system to run without internet connectivity.
- **Error Highlighting** – Visual indication of mispronounced words or phonemes.
- **Progress Tracking** – Recording and visualizing learners' improvements over time.
- **Accent Variation** – Differences in pronunciation patterns due to a learner's native language or dialect.
- **Phonetic Dictionary** – A mapping of words to their phoneme sequences, used for scoring and feedback.
- **Orthographic Dictionary** - A dataset with correct ortographic of a language.
- **Interactive Feedback** – Guidance that not only shows errors but suggests corrective actions, e.g., tongue placement or repetition prompts.
- **Vosk** – An offline speech-to-text engine, suitable for desktop use with moderate accuracy.
- **Whisper (OpenAI)** – A high-accuracy speech-to-text model, typically requires more computing resources.
- **Montreal Forced Alignment (MFA)** – A tool that aligns audio recordings with phonetic transcriptions, useful for analyzing precise pronunciation.
- **Phonetic Scoring Algorithm** – Any method that compares learner speech to target phonemes to produce a pronunciation score.
- **Audio Playback Module** – Component that allows learners to listen to their recorded speech for self-assessment.

2.1.3. Domain Terminology vs Rough Sketch

This section maps key terms to the raw observations and brainstorming notes collected in the Domain Rough Sketch. It demonstrates how the terminology was derived from real-world user and domain insights.

- **Learner Derived from:** Observations that users practicing English pronunciation are non-native speakers needing guidance.
- **Pronunciation Feedback Derived from:** Notes that learners require actionable feedback rather than binary correct/incorrect judgments.
- **Phoneme Derived from:** Observation that specific sounds (e.g., "th," "r/l," vowel contrasts) cause most learner errors.
- **Word-Level Accuracy Derived from:** Need to measure correctness for individual words in sentences for detailed progress tracking.
- **Speech-to-Text (STT) Engine, Vosk, Whisper (OpenAI) Derived from:** Research into available speech recognition technologies, evaluating accuracy, offline capability, and computational requirements.

- **Offline Mode Derived from:** Learner need to practice without continuous internet access.
- **Error Highlighting Derived from:** Observations that learners benefit from seeing which words or phonemes are mispronounced visually.
- **Progress Tracking Derived from:** Notes emphasizing motivation and monitoring improvement over time.
- **Accent Variation Derived from:** Observation that learners' native languages and accents influence pronunciation errors.
- **Phonetic Dictionary Derived from:** Research on tools like MFA and phonetic scoring methods to map words to their phonemes.
- **User Interface (UI) & Audio Playback Module Derived from:** Observations that learners need intuitive interfaces for recording, playback, and comparison.
- **Interactive Feedback & Phonetic Scoring Algorithm Derived from:** Notes that actionable guidance is more effective than numeric scores alone; requires phoneme-level scoring.
- **MFA (Montreal Forced Alignment) Derived from:** Research showing alignment tools improve the accuracy of phoneme-level analysis.

2.1.4. Narrative

In the modern landscape of language learning, many learners strive to improve their spoken English independently. While apps and courses provide vocabulary and grammar exercises, most learners struggle to obtain detailed feedback on pronunciation. Mispronunciations, especially of certain consonants, vowels, and clusters, often persist because learners lack immediate, actionable guidance.

Learners commonly attempt to self-correct by listening to recordings of native speakers or repeating phrases in apps. However, these methods provide limited insight, and without expert guidance, mistakes can be reinforced. Accent variation further complicates learning, as errors differ depending on a learner's native language.

Existing speech-to-text engines offer high transcription accuracy, but most are not optimized for pronunciation evaluation. Offline tools are rare, and online solutions may be expensive or require continuous connectivity. As a result, learners seeking independence and affordability often face barriers in effectively practicing pronunciation.

The domain narrative highlights a clear need: tools that empower learners to practice pronunciation accurately, monitor their own progress, and receive understandable, actionable feedback. Such tools would bridge the gap between the learner's effort and effective improvement, providing a path toward mastery without reliance on constant teacher intervention.

2.1.5. Events, Actions, Behaviors

This section categorizes key phenomena in the pronunciation coaching domain into **events**, **actions**, and **behaviors**:

Events: * Learner records a spoken sentence or word. * STT engine transcribes the spoken input. * System highlights mispronounced words or phonemes. * Learner receives a score or visual feedback on pronunciation accuracy. * Learner reviews progress dashboards or charts.

Actions: * Learner repeats a word or sentence to correct mispronunciation. * Learner listens to playback of their own pronunciation. * Learner consults phonetic hints or tips. * Learner tracks improvements over time using progress indicators.

Behaviors: * STT engine analyzes audio and use MFA and a algorthm to generates phoneme-level scoring. * Feedback module highlights errors and provides suggestions. * Progress tracking module updates visualizations and historical data. * Accent-aware algorithms adjust evaluation thresholds based on learner's background. * Offline mode ensures functionality without internet connectivity.

By separating these elements, the domain model clarifies **how the learner interacts with the domain** and **what the system must be able to observe or respond to**.

2.1.6. Function Signatures

The following functions describe operations in the pronunciation coaching domain. They are **conceptual and domain-focused**, grounded in what has been implemented or explored through research and prototyping.

Implemented Functions These functions have been actually implemented and tested:

- `recordSpeech(learnerInput)` → Captures the learner's spoken input as an audio recording.
- `transcribeSpeech(audio)` → Converts spoken audio into a textual transcription using the STT engines explored (Whisper, Vosk, Flutter libraries).

Explored / Researched Functions These functions have been studied, prototyped, or conceptually investigated but not yet implemented:

- `highlightErrors(transcription, target)` → Conceptually identifies mispronounced words or phonemes and generates visual or textual feedback.
- `computePhonemeScore(transcription, target)` → Investigated methods to calculate pronunciation accuracy at the phoneme level (e.g., using MFA or phonetic dictionaries).
- `playbackAudio(audio)` → Explored as a learner tool to listen to their recorded speech for self-assessment.
- `updateProgress(learner, score)` → Conceptually tracks and updates learner performance over time.
- `visualizeProgress(learnerData)` → Studied dashboards and visual representations to highlight trends, frequent errors, and improvement.
- `provideHints(mispronouncedPhonemes)` → Investigated ways to give actionable corrective suggestions.
- `exportProgressReport(learnerData)` → Considered exporting summaries of learner performance for personal review or tutor use.

Note: All explored functions are **derived from domain research and observations** and will be formally implemented in subsequent milestones.

2.2. Requirements

2.2.1. User Stories, Epics, Features

The following user stories capture the key functionality and goals of the Pronunciation Coach, based on implemented and explored features:

Epic 1: Recording and Transcription - **User Story 1.1:** As a learner, I want to record my spoken words and sentences so that I can practice pronunciation. - **Feature:** `recordSpeech()` function captures learner input. - **User Story 1.2:** As a learner, I want my speech to be transcribed into text so that I can see and confirmed what I said. - **Feature:** `transcribeSpeech()` function uses STT engines (Whisper, Vosk, Flutter libraries).

Epic 2: Feedback and Error Identification (Explored) - **User Story 2.1:** As a learner, I want mispronounced words or phonemes highlighted so that I know what to improve. - **Feature:** `highlightErrors()` (conceptually explored). - **User Story 2.2:** As a learner, I want to see a phoneme-level score for my pronunciation so that I can track accuracy. - **Feature:** `computePhonemeScore()` (explored through research).

Epic 3: Practice Support (Explored) - **User Story 3.1:** As a learner, I want to listen to my own recordings so that I can self-assess my pronunciation. - **Feature:** `playbackAudio()` (researched/prototyped). - **User Story 3.2:** As a learner, I want my progress tracked over time so that I can see improvement. - **Feature:** `updateProgress()` and `visualizeProgress()` (conceptually explored).

2.2.2. Personas

The following personas represent typical users of the Pronunciation Coach, highlighting their goals, challenges, and behaviors:

Persona 1: Ana – University Student - **Age:** 20 - **Background:** Non-native English speaker, studying at university in Puerto Rico. - **Goals:** Improve English pronunciation for presentations and exams. - **Challenges:** Limited time, struggles with certain vowel and consonant sounds, inconsistent feedback from existing apps. - **Behavior:** Practices pronunciation independently using apps, repeats phrases, and listens to recordings. - **Needs:** Immediate feedback, clear progress tracking, offline access.

Persona 2: Luis – Young Professional - **Age:** 28 - **Background:** Non-native English speaker, works in an international company. - **Goals:** Communicate clearly in meetings and calls, reduce accent-related misunderstandings. - **Challenges:** Limited opportunities for live feedback, difficulty identifying specific phoneme errors. - **Behavior:** Records himself speaking, compares to native pronunciation, uses feedback tools sparingly. - **Needs:** Accurate phoneme-level feedback, playback of recordings, easy-to-use interface.

Persona 3: Sofia – Language Enthusiast - **Age:** 16 - **Background:** High school student interested in learning English beyond school curriculum. - **Goals:** Speak English fluently for travel and online interactions. - **Challenges:** Motivation fluctuates, difficulty tracking improvement over time. - **Behavior:** Uses apps casually, likes interactive tools, occasionally seeks guidance from teachers or online communities. - **Needs:** Engaging feedback, progress visualization, ability to practice anytime.

Persona 4: Carlos – Mid-Career Professional - **Age:** 40 - **Background:** Non-native English speaker, currently working in a local company and seeking an international job opportunity. - **Goals:** Improve English pronunciation to communicate effectively in interviews and professional settings. - **Challenges:** Limited time for practice, anxiety about making mistakes, difficulty identifying specific pronunciation errors. - **Behavior:** Practices sporadically, prefers structured feedback, often listens to recordings to self-assess. - **Needs:** Accurate feedback on mispronunciations, clear guidance on improvement, progress tracking to stay motivated, flexible access (offline capability is important).

Summary: These personas represent the diversity of learners in terms of age, goals, and contexts. They help justify the **user stories, features, and design choices** made in this project.

2.2.3. Domain Requirements

The following domain requirements describe essential capabilities and constraints derived from the Pronunciation Coach domain, research, and user personas:

- **DR1 – Audio Capture:** The system must allow learners to record their spoken words or sentences accurately.
  - Justification: Ana and Luis need to practice pronunciation independently.
- **DR2 – Speech Transcription:** The system must convert learner speech into textual representation.
  - Justification: Provides learners with immediate feedback on what was spoken.
- **DR3 – Phoneme-Level Analysis:** The system should support evaluation of pronunciation at the phoneme level.
  - Justification: Mispronunciations often occur at specific sounds, which is critical for accurate feedback (all personas).
- **DR4 – Error Highlighting:** The system should indicate mispronounced words or phonemes to the learner.
  - Justification: Learners benefit from clear, actionable feedback.
- **DR5 – Progress Tracking:** The system should allow tracking of learner performance over time.
  - Justification: Learners like Ana and Sofia need motivation and insight into improvement.
- **DR6 – Playback Functionality:** The system should allow learners to listen to their own recordings.
  - Justification: Reinforces self-assessment and correction strategies.
- **DR7 – Accent Awareness:** The system should account for accent variations to improve feedback accuracy.
  - Justification: Luis and other learners with different native languages need reliable evaluation.
- **DR8 – Offline Operation:** The system should function without requiring continuous internet access.
  - Justification: Some learners may practice in environments with limited connectivity.

- **DR9 – Usability and Accessibility:** The system should have an intuitive interface suitable for learners of varying ages and technical proficiency.
  - Justification: Personas span ages 16–40 and different backgrounds.

Note: Additional requirements related to hints, export reports, or advanced AI feedback are **planned for future milestones** and are not included here as they have not been fully explored or prototyped.

2.2.4. Interface Requirements

The following interface requirements describe the interactions between the Pronunciation Coach system and its environment (learners, audio devices, and external resources):

- **IR1 – Audio Input Interface:** The system must accept audio input from the learner via microphone.
  - Source: Learner speaking into the device.
  - Observed phenomenon: Learner initiates a recording session.
- **IR2 – Audio Output Interface:** The system must provide audio playback of recorded speech.
  - Target: Learner listens to their own recordings.
  - Observed phenomenon: Learner plays back audio for self-assessment.
- **IR3 – Textual Output Interface:** The system must display transcription of spoken words and phoneme-level feedback.
  - Target: Learner sees text and error highlights.
  - Observed phenomenon: Learner reads transcription and evaluates pronunciation errors.
- **IR4 – Progress Visualization Interface:** The system should present graphical or tabular representations of learner performance over time.
  - Target: Learner monitors improvement trends.
  - Observed phenomenon: Learner reviews charts, scores, or dashboards.
- **IR5 – Accent-Aware Evaluation Interface:** The system should adjust feedback based on learner’s accent or native language.
  - Target: Learner receives personalized feedback.
  - Observed phenomenon: System analyzes speech patterns relative to accent.
- **IR6 – Offline Operation Interface:** The system should function without continuous internet access, handling both audio input and output locally.
  - Source/Target: Learner device.
  - Observed phenomenon: Learner interacts with system in offline mode.

Note: Additional interfaces for advanced AI feedback, hints, or report export are **planned for future milestones** and are not included here, as they have not been fully explored or prototyped.

2.2.5. Machine Requirements

The following machine requirements define technical constraints and performance expectations for the Pronunciation Coach system:

- **MR1 – Real-Time Audio Processing:** The system should process audio input and provide transcription within a maximum latency of 5 seconds for typical user recordings.
  - Justification: Ensures feedback feels fast and supports effective practice.
- **MR2 – Resource Usage:** The system should run efficiently on typical consumer devices (laptops, tablets, or smartphones) without excessive CPU or memory usage.
  - Justification: Ensures usability across a range of devices and prevents system slowdowns.
- **MR3 – Storage Requirements:** The system must store learner recordings and progress data efficiently, with each audio file ≤ 5 MB and overall user data ≤ 500 MB.
  - Justification: Maintains local storage limits while supporting offline operation.
- **MR4 – Accuracy Constraints:** The STT engine should achieve at least 85% transcription accuracy for standard learner speech in controlled testing scenarios.
  - Justification: Provides reliable feedback for learners; based on exploratory testing of Whisper and Vosk.
- **MR5 – Reliability and Stability:** The system should maintain operational stability during extended use (minimum 1-hour session) without crashes or data loss.
  - Justification: Ensures learner confidence and uninterrupted practice.
- **MR6 – Offline Capability:** The system must perform core functions (audio recording, transcription, playback) without internet access.
  - Justification: Supports learners practicing in environments with limited connectivity.

Note: Advanced requirements for AI hints or export functionality are **planned for future milestones** and are not included here, as they have not yet been prototyped or researched.

2.3. Implementation

2.3.1. Selected Fragments of Implementation

The following fragments illustrate the current implementation of the Pronunciation Coach system, highlighting architecture, user interface sketches, and code snippets for clarity.

Architecture Overview - The system follows a modular architecture with four primary components: 1. **Audio Capture Module** – Handles recording of learner speech from the microphone (`recordSpeech()` function). 2. **Speech-to-Text Module** – Converts recorded audio into text using multiple STT engines (`transcribeSpeech()` function), including Whisper, Vosk, and Flutter libraries for experimentation. 3. **Phoneme Analysis** - Take the text and use MFA to align the phoneme. Then, use an algorithm to evaluate the user's erros. 4. **Feedback and Visualization Module** – Responsible for providing error highlights, playback, and progress visualization (currently explored/researched, not fully implemented).

```
digraph G {
    rankdir=LR;
    AudioCapture -> SpeechToText -> FeedbackVisualization;
    AudioCapture [label="Audio Capture\n(recordSpeech)"];
    SpeechToText [label="Speech-to-Text\n(transcribeSpeech)"];
    FeedbackVisualization [label="Feedback & Visualization\n(highlightErrors,
playbackAudio, updateProgress)"];
}
```

Screen Sketches - **Recording Interface:** Simple button to start/stop recording, displays current session status. - **Transcription Display:** Text area showing learner's spoken words, with potential highlights for errors (conceptual). - **Playback Control:** Play, pause, and stop buttons for listening to recorded audio. - **Progress Visualization (Explored):** Prototype charts showing learner improvement over time.

3. Analytic Part

3.1. Concept Analysis

The Concept Analysis links the observations, research, and domain understanding collected in the rough sketch to the abstractions and terminology used in the Pronunciation Coach.

Rough Sketch → Abstractions - Observations from user behavior (recording, playback, practicing pronunciation) were abstracted into **core domain operations**: `recordSpeech()`, `transcribeSpeech()`, `highlightErrors()`, `computePhonemeScore()`, and `updateProgress()`. - Common patterns such as **mispronunciation detection** and **progress tracking** were identified as central concepts.

Abstractions → Terminology - The abstractions were then formalized into **domain-specific terms**: - **Learner**: the user practicing pronunciation - **Phoneme**: smallest distinguishable unit of sound - **STT Engine**: speech-to-text system used for transcription - **Feedback Module**: component providing error highlights and visual guidance - **Progress Visualization**: representation of learner improvement over time

Terminology → Narrative - The terminology was then incorporated into a **cohesive narrative** describing the learner's experience: - Learners record speech → system transcribes → errors are identified → learners receive feedback → progress is tracked over time. - This narrative captures the **flow of interactions and key concepts** independent of implementation, while grounding it in research and explored features.

Insights - Concept analysis demonstrates that all major domain concepts stem from **observed user needs and exploratory research**, ensuring that the system's design is grounded in reality. - It also highlights gaps where future work can extend functionality (e.g., hints, detailed reports, advanced AI feedback) without altering the core abstractions already explored.

Note: This analysis validates that the project's scope, features, and terminology are consistent with the **learners' needs and domain observations**, providing a solid foundation for implementation in subsequent milestones.

3.2. Validation and Verification

This section outlines the planned strategies for validating and verifying the Pronunciation Coach system against the requirements, user stories, and domain analysis.

Validation Approach - **Objective:** Ensure that implemented and explored features address learner needs effectively. - **Techniques:** - **Walkthroughs:** Team members simulate user interactions (recording, transcription, playback) to verify correct flow and usability. - **Scenario-Based Testing:** Test core functions with representative personas (Ana, Luis, Sofia, Carlos) to validate that transcription and recording work as intended. - **Cross-Accent Evaluation:** Test STT engines with different accents to verify transcription accuracy and reliability of feedback for diverse learners.

Verification Approach - **Objective:** Confirm that the system behaves as specified in requirements. - **Techniques:** - **Unit Testing:** Verify individual functions (`recordSpeech()`, `transcribeSpeech()`) for correctness. - **Integration Testing (Planned):** Assess interaction between Audio Capture and STT modules. - **Explored Modules Review:** Conceptual verification of feedback and progress visualization methods, ensuring design aligns with domain requirements.

Metrics for Evaluation - **Transcription Accuracy:** Measure percentage of correctly transcribed words compared to a reference. - **Latency:** Time between recording and transcription should be ≤ 2 seconds. - **Usability Feedback:** Collect qualitative feedback from team simulations or small pilot tests regarding interface clarity and learner experience.

Note: Full validation of explored modules (feedback, progress visualization, hints) will occur in subsequent milestones once prototypes or implementations are available.