

System Test Plan

For

Speech Recognition for Air Traffic Control

Team members: Braeden Burnett, Jakob Haehre, Kira McFadden, Tyler Carr

Version/Author	Date
1 All	9/17/22
1.1/All	10/28/22
1.2/All	11/1/22

Table of Contents

1. Introduction	1
1.1. Purpose	2
1.2. Objectives	2
2. Functional Scope	2
3. Overall Strategy and Approach	2
3.1. Testing Strategy	2
3.2. System Testing Entrance Criteria	2
3.3. Testing Types	3
3.3.1. Usability Testing	3
3.3.2. Functional Testing	3
3.4. Suspension Criteria and Resumption Requirements	3
3.4.1. Suspension Criteria	3
3.4.2. Resumption Requirements	3
Neural Model Resumption	3
4. Execution Plan	4
4.1. Execution Plan	4
5. Traceability Matrix & Defect Tracking	5
5.1. Traceability Matrix	5
5.2. Defect Severity Definitions	5
6. Environment	5
6.1. Testing Tools	5
6.2. Testing Environment	6
7. Assumptions	6
8. Risks and Contingencies	6
9. Appendices	6

1. Introduction

1.1. Purpose

This document is a test plan for Speech Recognition for Air Traffic Control System Testing, produced by the Scrum team. It describes the testing strategy and approach to testing the team will use to verify that the application meets the established requirements of the customer prior to release. It contains separate details for how to test both the website and the neural model. If these tests are completed, they validate that the system is complete. The system is not considered to be complete until all of the tests run successfully.

1.2. Objectives

Features that will be the objects of testing that are identified and classified into a hierarchy.

- Meets the requirements, specifications and the Business rules.
- Supports the intended business functions and achieves the required standards.
- Satisfies the Entrance Criteria for User Acceptance Testing.

2. Functional Scope

The Modules in the scope of testing for the Speech Recognition for Air Traffic Control System Testing are mentioned in the document attached in the following path:

/testing/README.md

3. Overall Strategy and Approach

3.1. Testing Strategy

Speech Recognition for Air Traffic Control System Testing will include testing of all functionalities that are in scope (Refer to Functional pe Section) identified. System testing activities will include the testing of new functionalities, modified functionalities, screen level validations, workflows, functionality access, testing of internal & external interfaces.

Website:

The website will be tested using Puppeteer to validate the presence of required elements on the page. Test cases will be written to navigate through the website and assert that the required items are present. The test cases will be generated from the requirements document.

Neural Model:

The neural model will be tested using prepared labeled testing data made from the provided ATC data. Using this data, we will calculate word error rate and word error rate per utterance for each of the models. We will be targeting a word error rate and word error rate per utterance of around ~.20. The current best word error rate claimed in industry is around ~.05[1]. However, this is achieved using the resources accessible by large corporations and thus is not within the scope of our project.

3.2. System Testing Entrance Criteria

In order to start system testing, certain requirements must be met for testing readiness. Readiness can be classified into:

Website:

The website will be ready for testing when all components that are required from the backlog to be included are present. The website must be hosted on the desktop machine provided and publicly available with a static IP address. Both versions of the maps must be functional with planes visible. The neural model must

be fully functional, and the website must be connected to the neural model.

Neural model:

The entrance criteria for the neural model testing is that the model(s) have been trained on the labeled ATC data. The testing scripts have been prepared and debugged and novel training data has been prepared.

3.3. Testing Types

3.3.1. Usability Testing

User interface attributes, cosmetic presentation and content will be tested for accuracy and general usability. The goal of Usability Testing is to ensure that the User Interface is comfortable to use and provides the user with consistent and appropriate access and navigation through the functions of the application (e.g., access keys, consistent tab order, readable fonts etc.)

3.3.2. Functional Testing

The objective of this test is to ensure that each element of the component meets the functional requirements of the customer as outlined in the:

- Functional Requirements
- Business rules or conditions
- Issue resolutions
- Feedback

3.4. Suspension Criteria and Resumption Requirements

This section specifies the criteria that will be used to suspend all or a portion of the testing activities on the items associated with this test plan.

3.4.1. Suspension Criteria

Testing will be suspended if the incidents found will not allow further testing of the system/application under-test. If testing is halted, and changes are made to the hardware, software or database, it is up to the Testing Manager to determine whether the test plan will be re-executed, or part of the plan will be re-executed.

Website Incidents:

- Website goes offline during testing
- Assertion test failed
- Unit test encounters exception

Neural Model Incidents:

- Model fails to load
- Training data is not novel to the model

3.4.2. Resumption Requirements

Resumption of testing will be possible when the functionality that caused the suspension of testing has been retested successfully.

Neural Model Resumption

- Check for issues in model file e.g., corrupted file, check all dependencies installed if no issues found attempt to reload model
- Get data novel to the model either from provided datasets or other methods

4. Execution Plan

4.1. Execution Plan

The execution plan will detail the test cases to be executed. The Execution plan will be put together to ensure that all the requirements are covered. The execution plan will be designed to accommodate some changes, if necessary, if testing is incomplete on any day. All the test cases of the projects under test in this release are arranged in a logical order depending upon their inter dependency.

Test Cases

ID	Name	System	Actions	Success Criteria
W1	Website load	Website	Load website	The website shall load without any errors and display the interactive map with a toggle to switch to the other map.
W2	Interactive map	Website	Load website	The map iframe is active.
W3	Map buttons zoom	Website	Click map zoom in button	The map iframe zooms in. The screenshot of the website before and after the zoom button pressed are different.
W4	Map pan	Website	Click on the map and drag 50 pixels to the right	The map iframe moves. The screenshot of the website before and after the zoom button pressed are different.
W5	Plane movement	Website	Load website	The planes on the map are updated with new coordinates as new data comes in. The screenshot of the website before and after a wait of 10 seconds is different.
W6	Click plane for details	Website	Locate a plane and click on it	An information panel is shown after clicking on the plane. The sections are checked to ensure the fields are filled in with plane information.
W7	Live transcription	Website	Locate a plane and click on it	The textbox is checked to ensure text is present. The live transcription is updated every 5 seconds, so the text is different after 5 seconds.
W8	Toggle button	Website	Click toggle button	The map is toggled between the interactive map and the sectional map. This is validated by checking that the classes are set to display none or display block.
M1	Word error rate	ASR Model	Run testing scripts for word error rate	The model achieves a word error rate of .2 or less
M2	Word error rate per utterance	ASR Model	Run testing scripts for word error rate per utterance	The model achieves a word error rate of .2 or less
M3	Precision, Recall, F1 scores	ASR Model	Run testing scripts for	The model achieves scores in those

			precision, recall, f1	
--	--	--	-----------------------	--

5. Traceability Matrix & Defect Tracking

5.1. Traceability Matrix

Requirements that start with an “A” indicate advanced requirements. These are not going to be addressed until a later sprint.

Requirement	Test Case
SIR 9	W7
SIR 10	W7
SIR 11	W1, W2, W5
SIR 12	W4
SIR 13	W5
SIR 14	W6
SIR 15	W7
USR 1	W8
USR 2	W6
USR 3	W6
USR 4	W3
USR 5	W4
AUSR 1	Coming soon
AUSR 2	Coming soon
AUSR 3	Coming soon
AUSR 4	Coming soon
AUSR 5	Coming soon
AUSR 6	Coming soon
AUSR 7	Coming soon
AUSR 8	Coming soon

5.2. Defect Severity Definitions

Critical	<ul style="list-style-type: none"> • API data is corrupted • Server is down • Model data is corrupted • Configuration Files from ATC Stream corrupted
Medium	<ul style="list-style-type: none"> • Plane loads into incorrect location • Text Box doesn't appear • Toggle Button doesn't work as anticipated • Word Error rate is higher than 0.2 for the model
Low	<ul style="list-style-type: none"> • Website map positioning loads incorrectly • Word Error rate is 0.2 for the model • Website takes longer than 5 seconds to load • Text Box loads incorrect text

6. Environment

6.1. Testing Tools

The System Testing Environment will be used for System Testing.

Website:

- Puppeteer

Neural Model:

- Manual validation of statistics

6.2. Testing Environment

The System Requirements for testing are as follows:

Hardware

- Dedicated Nvidia NeMo-compatible graphics card

Software

- Desktop: RedHat Enterprise Linux V8.6 or Ubuntu 22
- Laptop: Windows 11 Pro Version 21H2
- Python Version 3.6 to 3.8
- PyTorch Version 1.8.1

7. Assumptions

- Testers have access to a compatible NVIDIA GPU
- Puppeteer test cases are run using Node.js

8. Risks and Contingencies

Risk	Contingency
Plane coordinate fetching API not working	Display the last locations of the planes until the API comes back.
Model outputs poor transcription	Deploy an additional website for closed crowd-sourced model transcription validation.
Website is offline	Display an error message in place of the website indicating that the website is temporarily unavailable.
Model is unavailable	Temporarily pause transcriptions and replace with a message saying, “Temporarily unavailable”.

9. Appendices

a. Links

[1]

[https://smartaction.ai/blog/does-word-error-rate-matter/#:~:text=Word%20Error%20Rate%20\(WER\)%20is,word%20error%20rate%20of%204%25](https://smartaction.ai/blog/does-word-error-rate-matter/#:~:text=Word%20Error%20Rate%20(WER)%20is,word%20error%20rate%20of%204%25)