System Research & Test Plan

for

Kaldi Research Team

Prepared by
Adam Gallub, Milan Haruyama, Tabitha O'Malley, David Serfaty, Tahmina Tisha

Revision History

Name	Date	Reason For Changes	Version
All	10/18/2023	Write wrong information	V1.0
Tabitha	11/20/2023	Added in comments	V2.1
Tisha	11/27/2023	Sections 1, 1.2	V2.2
Tabitha	11/27/2023	Sections 3.3, 3.3.1, 3.3.2, 7, 8	V2.3
Milan	11/27/2023	Section 3.4; Editing all sections	V2.4
Tisha	11/28/2023	Section 3 Editing	V2.5
David	11/28/2023	Section 3.4, 8, 4.1.1, 4.1.2, 4.1.3, 1.2, 3.1.1, 3.3.1, 3.4.1, 3.4.2	V2.6
Tabitha	11/28/2023	Section 1.1 Editing/Reading All Sections Editing 3.3.2, 4, 5	V2.7
Adam	11/28/2023	Section 2, 3.2; Editing, Writing to the sections	V2.8
Milan	11/28/2023	Editing all sections	V2.9
Tabitha	11/29/2023	Section 3.1	V2.10
Tisha	11/29/2023	Section 3.; Edited the section	V2.11
Milan	12/03/2023	Editing all sections	V2.12
Milan	12/04/2023	Editing all sections	V2.13

Table of Contents

1. Introduction

1.1 Purpose

This document is a test plan for the Kaldi Research Team's testing phase for the RTube project. Outlined below are the testing strategies that the Kaldi Research Team shall use to verify the impact of modifications made to the existing automatic speech recognition (ASR) model before its implementation in the RTube web application.

1.2 Objectives

Objective 1	Input Testing	Assess the outcomes of all possible inputs.
Objective 2	Accuracy Testing	Assess the transcription accuracy of the ASR model using predefined benchmarks (e.g., a minimum transcription accuracy of 80%).
Objective 3	Runtime Testing	Evaluate the efficiency of the ASR model in terms of processing time. It shall measure the time taken by the ASR model to process a standard WAV file and produce the output text file.

2. Functional Scope

The modules in the testing scope for the Kaldi ASR Toolkit are outlined in the documents attached in the following path:

System Requirements Specification Document (Final Draft)	Section 3: Software Interface Requirements
	Section 4: System Features
System Design Document (Final Draft)	Section 3: Human-Machine Interface
	Section 5.2: Interface Detailed Design
User Manual	To Be Determined
System Research & Test Plan (Final Draft)	Section 3.1: Testing Strategy
	Section 3.3: Functional Testing
	Section 3.4: Suspension Criteria & Resumption Requirements
	Section 4: Execution Plan

3. Overall Strategy and Approach

3.1 Testing Strategy

Testing shall evaluate the ASR model's ability to process different audio input formats, its transcription accuracy, and its runtime speed.

3.1.1 Input Testing

Test Objective:
Test all possible inputs.

Technique:

- Input a MP3 file
- Use nothing as an input
- Input multiple files of any type
- Compare all errors and exceptions thrown
- Input a WAV file

Completion Criteria:
The system shall throw an error and/or exception upon reception of an input that does not meet the system requirements or successfully execute the program.

Special Consideration:
If the program terminates and no exception or error is thrown then this indicates that the Kaldi ASR Toolkit is not functioning properly and shall be fixed or reinstalled.

3.1.2 Accuracy Testing

Test Objective:
Verify the accuracy of the transcription generated by the ASR model.

Technique:

- Use a WAV file containing Air Traffic Control (ATC) transmission data to evaluate transcription accuracy.
- Compare the ASR model's transcriptions with the correct transcription.
- Calculate the transcription accuracy based on the percentage of correct words in the transcription.

Completion Criteria:
The system shall have a transcription accuracy of at least 80%.

Special Consideration:
If the accuracy falls below the benchmark, the ASR model shall be retrained and retested until transcription accuracy meets the benchmark.

The transcription accuracy shall be calculated by dividing the number of correct words by the total number of words in the transcription.

3.1.3 Runtime Testing

Test Objective:
Ensure that the ASR model generates a transcription within an acceptable runtime.

Technique:

- Measure the runtime of the ASR model.
- Establish a runtime benchmark that the ASR model must meet (e.g., five minutes).

Completion Criteria:
The system shall generate a transcription within the established runtime benchmark.

Special Consideration:
If the runtime exceeds the established benchmark, the ASR model shall be retrained to mitigate performance issues.

3.2 System Testing Entrance Criteria

Upon the availability of a dataset and a base model, the system shall be defined as ready. The dataset shall consist of 30 hours of training data used for accuracy testing (*Section 3.3.1*). The base model is the sample ASR model included with the Kaldi ASR toolkit whose runtime shall be tested (*Section 3.3.2*). The base model may be changed in the future.

3.3 Functional Testing

The functionality of the ASR model shall be tested once it is trained using the 30-hour ATC dataset provided by the product owner. More specifically, input reception, transcription accuracy, and runtime shall be tested. Upon being able to generate a sufficiently accurate transcription within an adequate runtime, the ASR model shall be implemented in the RTube web application in the later future. The implementation of the ASR model shall require its own set of functional testing that shall be described in future iterations of this project.

3.3.1 Input Testing

All possible inputs shall be tested to ensure that the system shall output a transcription upon receiving a valid input. Upon receiving an invalid input (e.g., a non-WAV file, multiple files of any type), the system shall throw an exception and/or error.

3.3.2 Accuracy Testing

A high transcription accuracy is critical due to the low margin of error required to communicate effectively with ATC. As such, any ASR model with a transcription accuracy lower than 80% shall be retrained and retested until the benchmark is met or surpassed.

3.3.3 Runtime Testing

A fast runtime is critical due to the high demand of the end goal of this project - transcribing live ATC transmissions in real time. As such, any ASR model with a runtime slower than five minutes shall be retrained and retested until the benchmark is met or surpassed.

3.4 Suspension Criteria & Resumption Requirements

This section shall detail the suspension criteria and resumption requirements used to determine how to proceed with testing upon reaching a plateau in improvement, or upon receiving any fatal errors during testing.

Suspension Criteria	Resumption Criteria
ASR model failing to meet performance benchmarks (e.g., transcription accuracy of at least 80%, maximum runtime of five minutes).	Creation of a new ASR model iteration.
ASR model failing to compile or execute.	Creation of a new ASR model iteration or reinstallation of current iteration.
Kaldi ASR Toolkit failing to compile or execute.	Reinstallation and reconfiguration of Kaldi, and creation of a new ASR model iteration.

4. Execution Plan

4.1 Execution Plan

This section shall detail the test cases used to ensure proper execution of the ASR model. Input testing is the precondition for execution; accuracy and runtime testing are the postconditions. Additional test cases shall be appended as needed.

4.1.1 Input Testing

Requirement	Test Case ID	Input	Expected Behavior	Pass/Fail
13.1.1 The system shall throw an exception upon a non-WAV file being inputted.	1.1	The system receives a non-WAV file at time equals 0.	The system shall throw an exception and alert the user that it was expecting a single WAV file as input.	Pass
13.1.2 The system shall throw an exception upon multiple files of any type being inputted.	1.2	The system receives multiple files of any type at time equals 0.	The system shall throw an exception and alert the user that it was expecting a single WAV file as input.	Pass
13.1.3 The system shall throw an exception upon no files being inputted.	1.3	The system receives no input at time equals 0.	The system shall throw an exception and attempt a traceback of the most recent WAV file in the directory.	Pass
13.1.4 The system shall output a text file called “out.txt” that contains the transcription of the inputted WAV file.	1.4	The system receives a WAV file at time equals 0.	The system shall output a text file called “out.txt” that contains the transcription of a WAV file.	Pass

4.1.2 Accuracy Testing

Requirement	Test Case ID	Input	Expected Behavior	Pass/Fail
14.1.1 The system shall output a text file called "out.txt" that contains a transcription with a minimum accuracy of 80% .	2.1	The system receives a WAV file composed of spoken English at time equals 0.	The system shall return the transcribed speech as a text file called "out.txt" with a minimum transcription accuracy of 80%.	Pass

4.1.3 Runtime Testing

Requirement	Test Case ID	Input	Expected Behavior	Pass/Fail
15.1.1 The system shall output a text file "out.txt" within five minutes of activation.	3.1	The system receives a WAV file composed of spoken English at time equals 0.	The system shall return the transcribed speech as a text file called "out.txt" within five minutes of activation.	Pass

5. Traceability Matrix & Defect Tracking

5.1 Traceability Matrix

This section contains a list of requirements and their corresponding test cases.

5.1.1 Critical Requirements

Requirement	Description`	Test Case
System Requirements Specification 13.1.1	The system shall throw an exception upon a non-WAV file being inputted.	Check if a non-WAV file has been inputted.
System Requirements Specification 13.1.2	The system shall throw an exception upon multiple files of any type being inputted.	Check if multiple files have been inserted.
System Requirements Specification 13.1.3	The system shall throw an exception upon having no files being inputted.	Check if no file was inserted.
System requirements Specification 13.1.4	The system shall output a text file "out.txt" that contains the transcription of the inputted WAV file.	Check if the inputted file was a WAV file.

5.1.2 Medium Requirements

Requirement	Description`	Test Case
System Requirements Specification 14.1.1	The system shall output a text file called "out.txt" that contains a transcription with a minimum accuracy of 80%	Check that the words contained in "out.txt" match the correctly transcribed WAV file.
System Requirements Specification 15.11	The system shall output a text file "out.txt" within five minutes of activation	Measure the runtime upon system execution.

5.2 Defect Severity Definitions

Critical	The defect causes a catastrophic failure that results in loss of functionality for the user. A manual procedure requiring a high level of effort must be implemented to remedy the defect, if at all possible to remedy in the first place. Examples of critical defects include: ● System abends ● Data cannot flow through a business function/lifecycle ● Data is corrupted or cannot post to the database
Medium	The defect does not seriously impair system function. A manual procedure requiring a medium level of effort can be implemented to remedy the defect. Examples of a medium defect include: ● Form navigation is incorrect ● Field labels are not consistent with global terminology
Low	The defect is cosmetic or has a negligible effect on system functionality. A manual procedure requiring a low level of effort can be implemented to remedy the defect. Examples of a low defect include: ● Repositioning of fields on screens ● Text font on reports is incorrect

6. Environment

- The testing environment is the Ubuntu command line via the Windows Subsystem for Linux (WSL).

7. Assumptions

- The dataset is sufficient to train an ASR model with a minimum transcription accuracy of 80% accuracy and maximum runtime of five minutes.
- The dataset used to train the ASR model shall be in General American English.
- The dataset used to train the ASR model shall include ATC vernacular.
- The text transcribed by the ASR model shall be based on General American spelling conventions (e.g., “color” versus “colour”).
- All testing performed on the ASR model shall be done under the assumption that the system can compile, train, and run the model.
- All tests shall be run to completion without interruption or changes partway through.
- The system shall retrain the model if the accuracy or runtime do not meet the requirements .

8. Risks and Contingencies

ID	Risk	Impact	Contingency Plan
1	Unable to decrease runtime.	Low	Upgrade GPU or install additional one.
2	Unable to increase accuracy.	Low	Retrain model.
3	Kaldi ASR Toolkit termination error.	High	Reinstall the Kaldi ASR Toolkit.
4	Unable to improve the model.	High	Retrain model.

9. Appendix: Glossary

Term	Definition
ATC	***Air Traffic Control;*** the service that elicits communications between pilots and helps to prevent air traffic accidents.
ASR	***Automatic Speech Recognition;*** the ability for computers to recognize and translate spoken speech.
CLI	***Command-Line Interface;*** text-based interface that allows interaction from the user to the computer program.
DNN	***Deep Neural Network;*** a machine learning technique that represents learning and processing data in artificial neural networks.
ERAU	***Embry Riddle Aeronautical University;*** an aviation-centered university located in Daytona Beach, Florida, United States.
FFmpeg	***Linux utility;*** a portable open-source utility that allows users to decode, encode, transcode, multiplex, demultiplex, stream, filter, and play most human- or machine-made multimedia.
FFT	***Fast Fourier Transform;*** algorithm used to obtain the spectrum or . frequency content of a signal.
General American English	The most spoken variety of the English language in the United States.
GMM	***Gaussian Mixture Model;*** used to calculate the distance between the MFC feature vector and the HMM state.
HMM	***Hidden Markov Model;*** used to find the state locations of the phonemes..
IPA	***International Phonetic Alphabet:*** an alphabetic system of phonetic notation developed by the International Phonetic Association; used to represent speech sounds in a standardized format.
Lexicon	*pertaining to speech;* a library of words that is understood by the language model.
MFC	***Mel-Frequency Cepstrum;*** a representation of the short-term power spectrum of a sound
MFCCs	***Mel-Frequency Cepstral Coefficients;*** the coefficients that a MFC is comprised of
NLP	***Natural Language Processing;*** the culmination of computer science, linguistics, and machine learning.
Phone	*pertaining to speech;* a distinct speech sound or gesture;
Phoneme	*pertaining to speech;* a set of phones that can distinguish one word from another
Triphone	*pertaining to speech;* a sequence of three consecutive phonemes
WER	***Word Error Rate;*** the rate at which error in words occurs
WSL	***Window Subsystem for Linux;*** allows users to run a GNU/Linux environment directly on Windows without the overhead of running the environment through a virtual machine.