

EchoText

User Manual
Version 1.00

Team: Ahmad Ghaddar, Kevin Xing, Harsh Bhagat, Andy Huang

Revision History

Date	Description	Author	Comments
12/3/24	Version 1.00	Andy Huang	First Revision

Document Approval

The following User Manual has been accepted and approved by the following:

Signature	Printed Name	Title	Date
	Kevin Xing	Team Lead Meeting Minutes	
	Ahmad Ghaddar	Presentation Lead Frontend Lead UI/UX Lead	
	Harsh Bhagat	GitHub Lead Backend and Data Management Lead	
	Andy Haung	Documentation Lead Testing/QA Lead	

Table of Contents

REVISION HISTORY	I
DOCUMENT APPROVAL	I
1. INTRODUCTION	1
1.1 PURPOSE OF THE APPLICATION	1
2. GETTING STARTED	1
2.1 REQUIREMENTS	1
2.2 INSTALLATION	1
3. UI/FUNCTIONALITY	2
3.1 MAIN PAGE VIEW	2
3.1.1 Header	2
3.1.2 Text Input Area	2
3.1.3 Action Buttons and Progress Bar (Bottom Section).....	3
3.1.4 Navigation Bar (Bottom Menu).....	3
3.2 GENERATED VIEW	4
3.1.1 Top Toolbar.....	4
3.1.2. List of Saved Audios	4
3.1.3 Playback Controls	5
3.1.4 Seek Bar.....	5
3.1.5 Context Menu.....	5
3.3 DOCUMENTS VIEW	6
3.1.1 Top Toolbar.....	6
3.1.2 List of Saved Documents	6
3.1.3 Import	6
3.1.4 Document selection	7
3.1.5 Page Selection	8
3.1.6 Context Menu.....	9
3.4 HISTORY VIEW	10
3.1.1 History Counter	10
3.1.2 Clear All	10
3.1.3 Copy and Delete.....	10
3.5 SETTINGS VIEW	11
3.5.1 Main Settings Buttons	11
3.5.2 Speech Settings	12
3.5.3 Features in the Theme & Display Section	12
3.5.3 Theme & Display.....	13
3.5.4 Features in the Theme & Display Section	13
3.5.5 Security.....	14
3.5.6 Features in the Security Section	14
3.5.4 ADVANCED	16
Features in the Advanced Section:.....	16
3.5.5 SUPPORT	17
Features in the Support Section	17
4. GETTING EXTERNAL MODELS	18
4.1 GETTING VITS/ONNX MODEL	18
4.2 IMPORTING A VITS MODEL TO ONNX USING PYTHON	18
5. SUPPORT	19

1. Introduction

Welcome to the EchoText User Manual. This guide provides detailed instructions for using the EchoText application, a transformer-based, on-device Text-to-Speech (TTS) solution designed for iOS. This User Manual is designed to help you understand and make the most of EchoText's features. It provides step-by-step instructions on how to set up and use the app, from voice profile customization to adjusting themes and managing security settings. Whether you're a new user or experienced with TTS apps, this manual will guide you through the functionality of EchoText.

1.1 Purpose of the Application

The purpose of the EchoText application is to convert text to speech. EchoText allows users to convert text into speech entirely offline, ensuring privacy, security, and performance. The app leverages the latest TTS technology with a focus on providing a seamless and intuitive experience on iPhone XR and higher devices. Whether you're reading documents, listening to articles, or using the app for accessibility purposes, EchoText offers a customizable voice experience, complete with adjustable pitch, speed, and voice model settings.

2. Getting Started

2.1 Requirements

Before you begin using EchoText, please ensure that your device meets the following requirements:

- Hardware Requirements:
 - iPhone XR or higher (supports the latest iOS versions)
 - Minimum free storage space: 2GB (for app installation and initial setup)
- Software Requirements:
 - iOS Version: iOS 17.0 or later

Other Requirements:

- Internet Connection
 - While the app works offline for generating speech, an internet connection is required for initial setup, downloading necessary components, and occasional updates.

2.2 Installation

Follow the steps below to install EchoText on your iPhone and get started.

- Step 1: Open the App Store on your iPhone.
- Step 2: Search for EchoText in the search bar.
- Step 3: Tap the Download button or “Get” icon to the right, to start the installation process.

- Step 4: Wait for the app to download and install on your device. This may take a few minutes depending on your internet speed.
- Step 5: Begin using EchoText for your Text-to-Speech needs.

3. UI/Functionality

3.1 Main Page View

This is the primary interface where you can input text and generate speech.

The Main Page is the initial screen you see when you launch EchoText, acting as the central hub for starting text-to-speech generation. You can easily return to the Main Page at any time by selecting the Main Tab (Home Icon) in the menu bar.

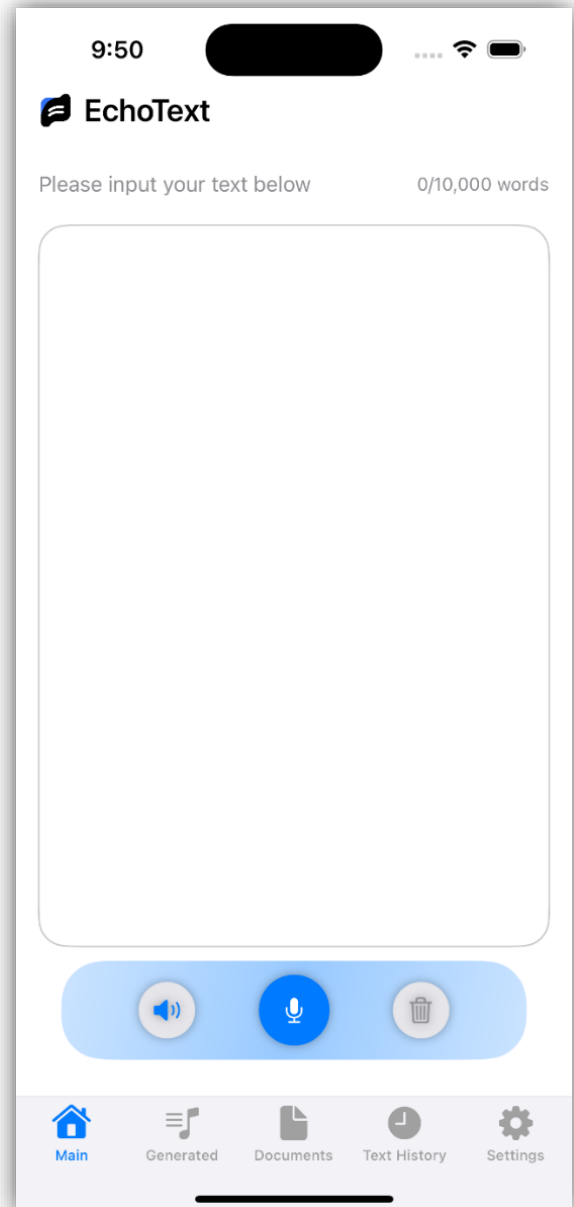
Below is a detailed explanation of the components and functionality of this page:

3.1.1 Header

- **Logo and App Name:**
 - Displays the app name ("EchoText") alongside the app logo for easy identification.
- **PDF counter:**
 - During batch document processing, a PDF counter will show on the far-right side that displays the current pdf being generated out of the total within the batch.

3.1.2 Text Input Area

- **"Please input your text below":**
 - Instructional text that prompts you to type or paste your desired text for text-to-speech generation.
- **Word Counter:**
 - Located in the top-right corner above the text box. Tracks the number of words entered, with a maximum limit of 10,000 words.
- **Text Input Box:**
 - The large box in the center of the screen where you can type or paste your text. This is the main area for entering content that will be converted into speech.

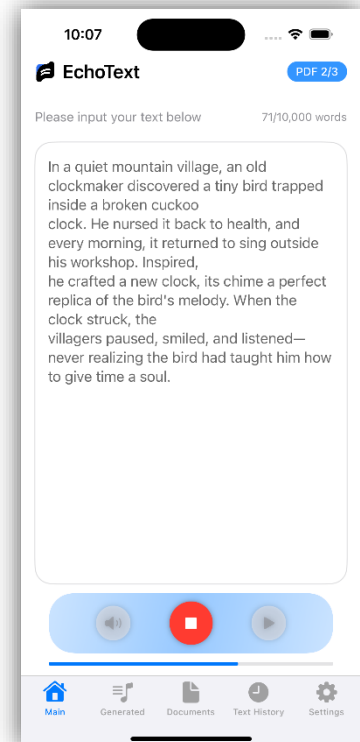
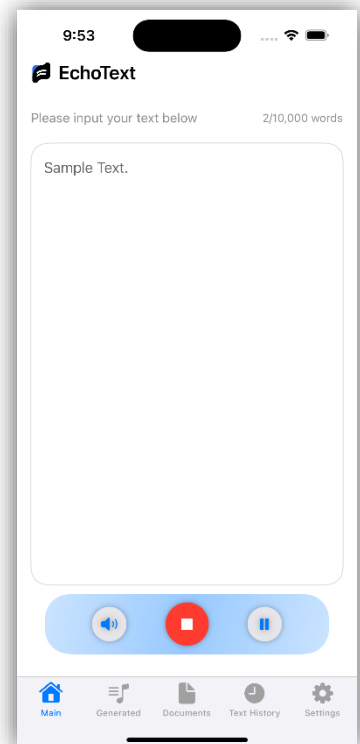


3.1.3 Action Buttons and Progress Bar (Bottom Section)

- **Speaker Icon (Mute/Unmute):**
 - Toggles the audio output of the current text-to-speech generation on or off.
 - Use this to quickly mute or unmute the playback during or after generation.
- **Microphone Icon (Generate/Terminate):**
 - Starts the text-to-speech generation process when pressed.
 - During generation, the microphone icon switches to a red circle with a square inside it, allowing you to terminate the ongoing process if needed.
- **Trash Icon (Clear Text) / Play-Pause:**
 - Clears all text from the input area when not generating.
 - During text-to-speech generation, this icon changes to a Play/Pause button, allowing you to play and pause the playback of the currently generating audio.
- **Progress Bar:**
 - Appears below the action buttons during text-to-speech generation.
 - Visually indicates the progress of the generation process, helping you monitor completion in real-time.

3.1.4 Navigation Bar (Bottom Menu)

- **Main:** The current page. It is marked with a blue home icon to indicate that you are on the main text input interface.
- **Generated:** Access previously generated audio files.
- **Documents:** Import and manage documents for generation or later use.
- **Text History:** Review your recent text inputs.
- **Settings:** Customize your experience and manage your preferences.



3.2 Generated View

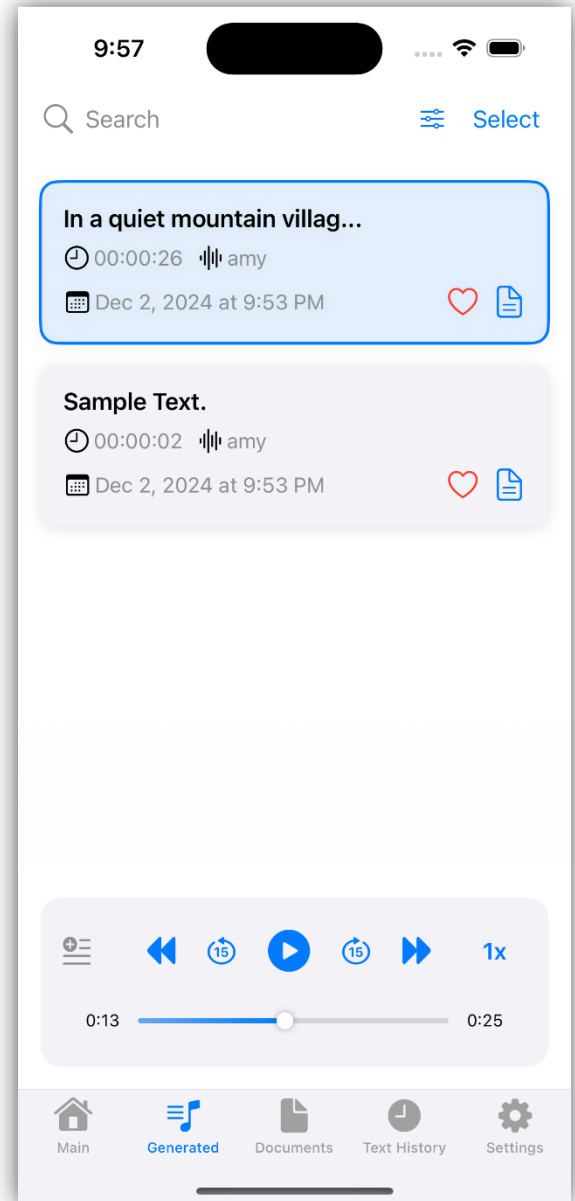
The Generated Page is where all your saved audio files are displayed and managed. This page provides a variety of tools for sorting, playback, and managing your generated audio files. Below is a breakdown of its features and functionality.

3.1.1 Top Toolbar

- **Search Icon:** Opens a search bar where you can quickly find specific audio files by name.
- **Filter Icon:** Opens the **Sorting and Filters Sheet**, where you can:
 - Sort by date, length, or name.
 - Filter audio files based on parameters like speed, pitch, length, favorites, includes document, date range and Voice Model.
- **Select Button:** Allows you to select multiple audio files for bulk actions like sharing or deleting.

3.1.2. List of Saved Audios

- Audio files are displayed in a **card-like style**, with the following information:
 - **Name:** The title of the audio file.
 - **Length:** Duration of the audio (e.g., 00:26).
 - **Date and Time:** When the audio was generated.
 - **Voice Model:** The name of the voice used for generation (e.g., "Amy").
- Each card includes:
 - **Heart Icon:** Tap to mark the audio as a favorite.
 - **Document Viewer Icon:** Tap to view the text used to generate the audio.



3.1.3 Playback Controls

Located below the audio list, the playback controls allow for seamless navigation and playback of your audio files. The controls include:

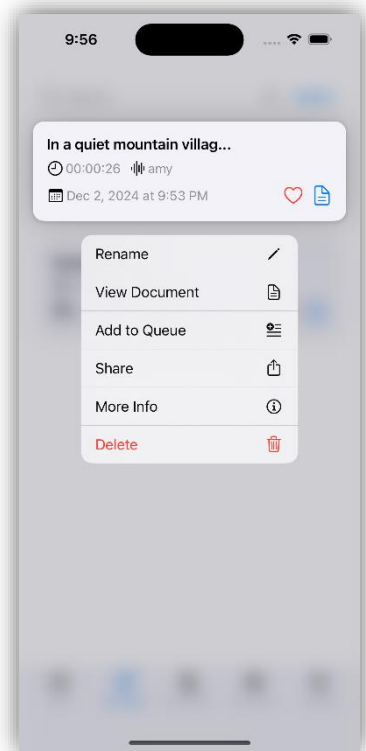
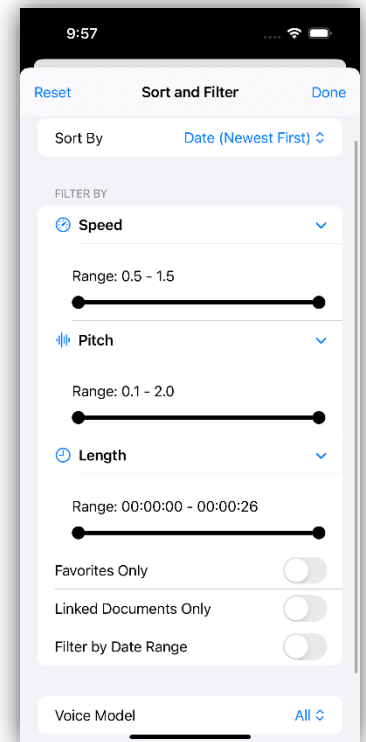
1. **Queue Icon:** Opens the queue view to manage upcoming audio playback.
2. **Previous Audio:** Skips to the previous audio file in the list.
3. **Rewind 15 Seconds:** Rewinds the playback by 15 seconds.
4. **Play/Pause:** Toggles between playing and pausing the current audio.
5. **Forward 15 Seconds:** Advances the playback by 15 seconds.
6. **Next Audio:** Skips to the next audio file in the list.
7. **Speed Control:** Allows you to adjust the playback speed (e.g., 1x, 1.5x, 2x).

3.1.4 Seek Bar

- **Seek Bar:** Displays the progress of the currently playing audio.
- **Current Time:** Shown on the left side of the seek bar, indicating the elapsed time.
- **Total Time:** Shown on the right side of the seek bar, indicating the total duration of the audio.

3.1.5 Context Menu

- **Accessing the Menu:** Hold down on any audio file to bring up the context menu.
- **Menu Options:**
 - **Rename:** Edit the name of the audio file.
 - **View Document:** Opens the text used for generating the audio.
 - **Add to Queue:** Adds the audio file to the playback queue.
 - **Share:** Allows you to share the audio file with others.
 - **More Info:** Provides detailed metadata about the audio file.
 - **Delete:** Permanently removes the audio file.



3.3 Documents View

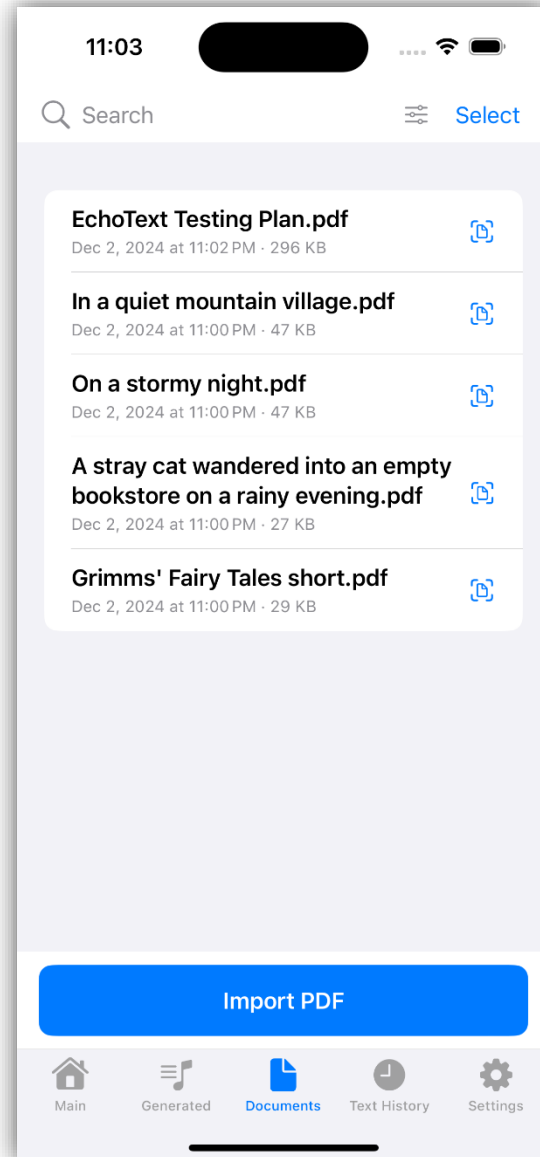
The Documents View is designed to help users manage, search, and interact with text-based documents for text-to-speech conversion. This view provides a user-friendly interface for importing, searching, and converting documents.

3.1.1 Top Toolbar

- **Search Icon:** Opens a search bar where you can quickly find specific audio files by name.
- **Filter Icon:** Opens the Sorting and Filters Sheet, where you can:
 - Sort by date, length, or name.
 - Filter audio files based on parameters like speed, pitch, length, favorites, includes document, date range and Voice Model.
- **Select Button:** Allows you to select multiple PDF files for bulk actions like deleting and batch conversion.

3.1.2 List of Saved Documents

- documents files are displayed in a **list-like style**, with the following information:
 - **Name:** The name of the imported PDF.
 - **Date and Time:** The date the PDF was imported.
 - **File Size:** the file size of the PDF.
- **Each PDF includes:**
 - **Document Viewer Icon:** Tap to view icon at the right side to view the PDF that was imported.

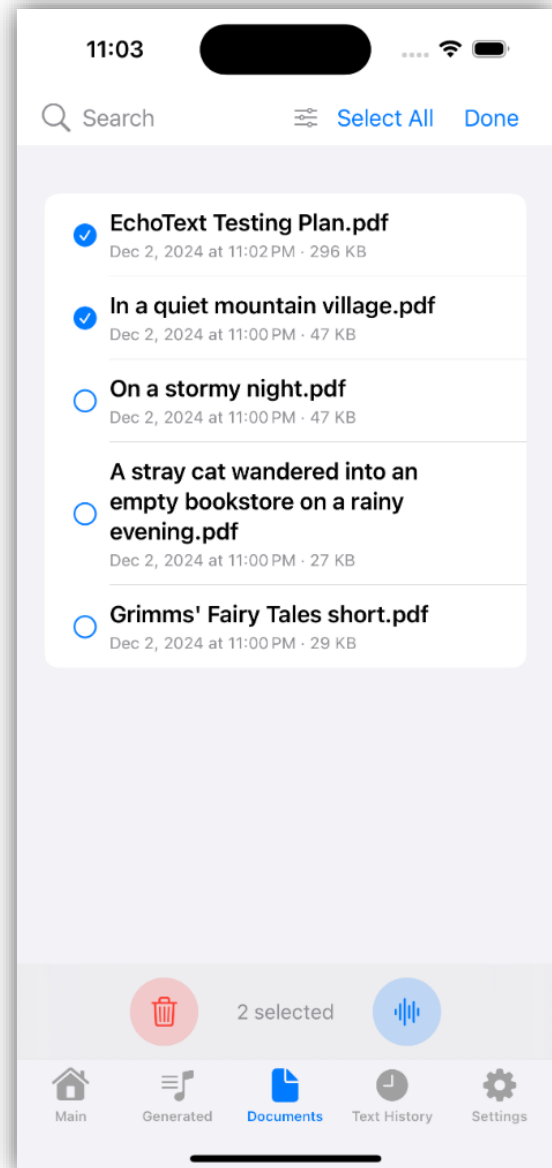


3.1.3 Import

At the bottom of the screen, a blue “Import PDF” is where you would press to import PDF documents for conversion. By pressing said button, it will pull up Apple Files app which you can select your desired PDF to import.

3.1.4 Document selection

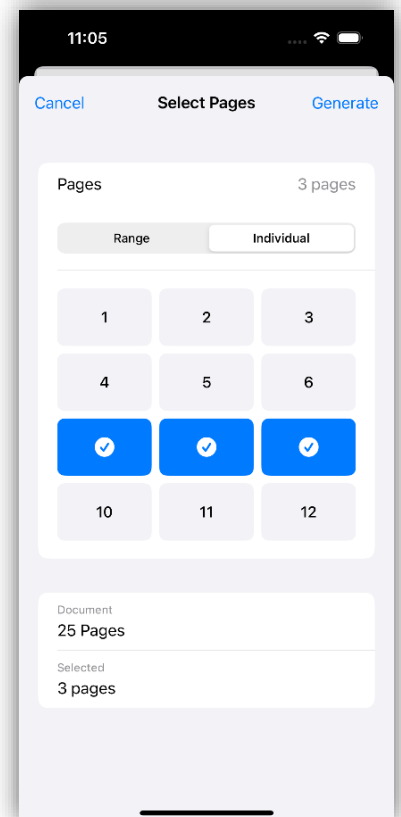
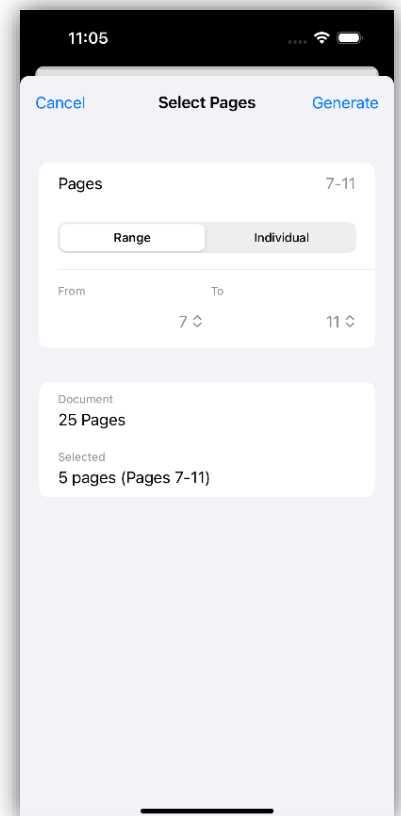
- **Select all:** Select all documents that are displayed to the user.
- **Done:** to exit out of selected mode.
- **Select document:** the user can press on the individual document to select. A blue icon with a checkmark will be displayed to let the user know that the document has been selected.
- **Delete all Selected:** By pressing on the red trash icon at the bottom of the screen, the user can delete all selected documents. A confirmation popup will show to ensure if the user wants to perform said action.
- **Batch Conversion:** By pressing on the blue wave icon at the bottom of the screen, the user can convert all the documents selected to audio. The app will queue each document to be converted displayed in Main View and the audio file will be saved and displayed in the Generated view.



3.1.5 Page Selection

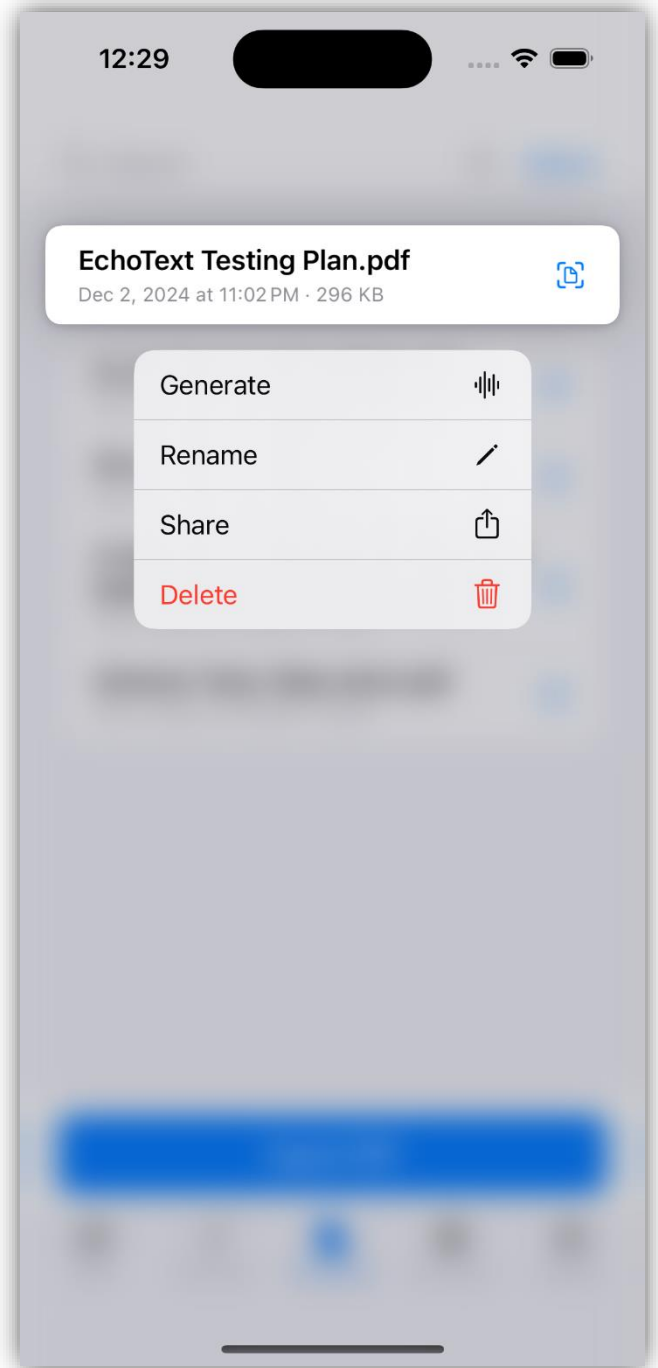
In the document preview page, the user can press on Select Pages button at the bottom of the app to open up a page selection page for conversion.

- **Select Range:** The user can select a page range to convert said pages to text. There is a dropdown menu next to the “From” text which indicate the start. There is another dropdown menu next to the “To” text which indicates the end. At the bottom, it shows the total number of pages, selected pages, and page range.
- **Select Individual:** The user can select individual pages displayed by a grid view in the middle of the app. The user can click on each box to select which is indicated by a blue box and checkmark. At the bottom show total pages and selected pages.
- **Cancel:** at the top right there is a cancel button to exit out of Select Page view.
- **Generate:** at the top left there is a Generate button to convert the pages to audio.



3.1.6 Context Menu

- **Accessing the Menu:** Press and hold down on any document file to bring up the context menu.
- **Menu Options:**
 - **Generate:** To convert the document to audio.
 - **Rename:** Edit the name of the document file.
 - **Share:** Allows you to share the document file with others.
 - **Delete:** Permanently removes the document file.



3.4 History View

The History View in EchoText provides a user-friendly interface for managing all your previously generated audio files. This feature ensures that users can quickly access, organize, and manage their TTS (Text-to-Speech) history. Whether you need to delete unnecessary files or copy files. History View offers seamless functionality for an optimal user experience. Below is a breakdown of its features and functionality.

3.1.1 History Counter

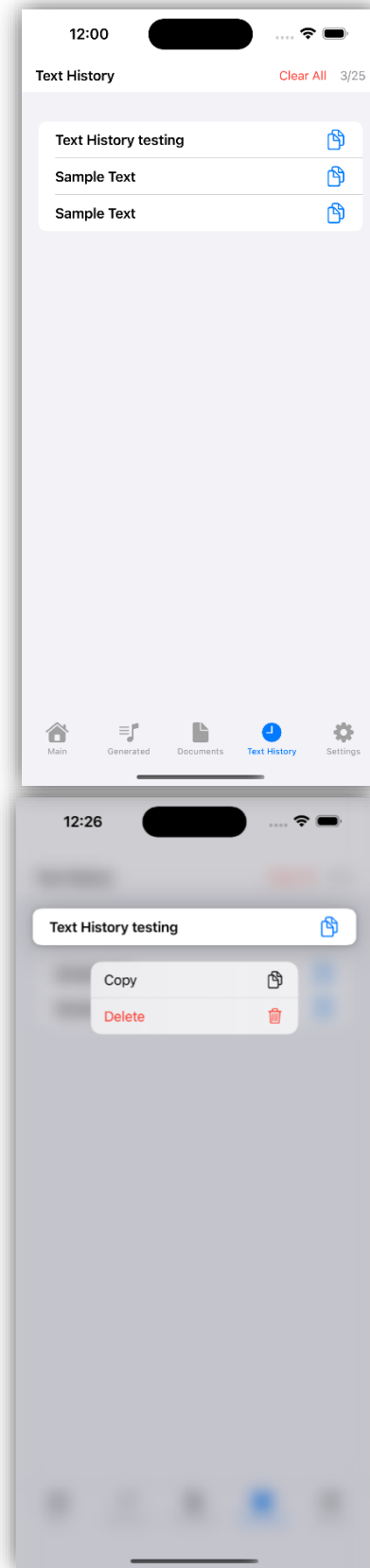
- The counter at the top right corner of the Text History displays the number of files currently in your history. It will save the last 25 generated audio history.

3.1.2 Clear All

- If you want to remove all generated audio files at once, the Clear All button provides a convenient option.
- Tap the Clear All button at the top right of the Text History to delete all files from your history. You will be prompted to confirm this action before the files are removed permanently.

3.1.3 Copy and Delete

- **Copy:** To share or save the audio file elsewhere, you can copy it. Simply tap the copy icon next to the file, and it will be copied to your clipboard, ready to be pasted wherever needed. Alternatively, you can press and hold the file to bring up the copy option.
- **Delete:** To remove a file from your text history you can press and hold the file to bring up the delete option.



3.5 Settings View

The Settings section in EchoText allows you to customize your experience and manage your preferences. The settings section is divided into five sections for easy navigation.

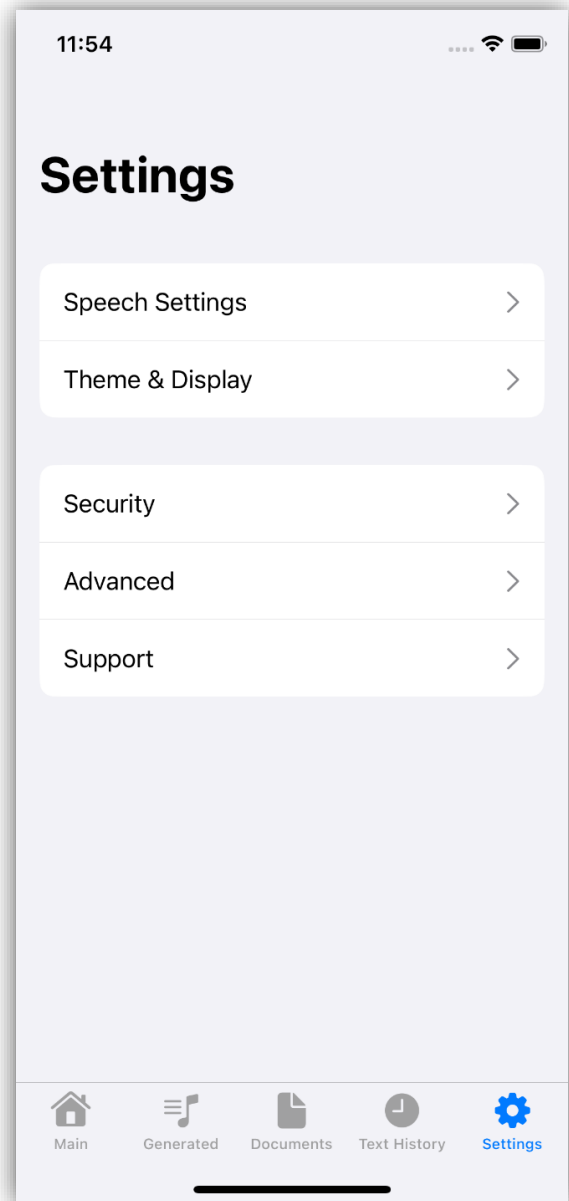
To access the settings page:

1. Open the application.
2. Tap the gear icon located in the bottom-right corner of the home screen.

3.5.1 Main Settings Buttons

The main settings page displays the following buttons:

Button	Description
Speech Settings	Adjust text-to-speech options, including voice selection, tone, pitch, and speed.
Theme & Display	Customize the app's appearance, including themes (light/dark) and font settings.
Security	Set up and manage passwords, encryption, and other security features.
Advanced	Access power-user configurations and experimental settings.
Support	Find help, FAQs, and contact information for technical support.



3.5.2 Speech Settings

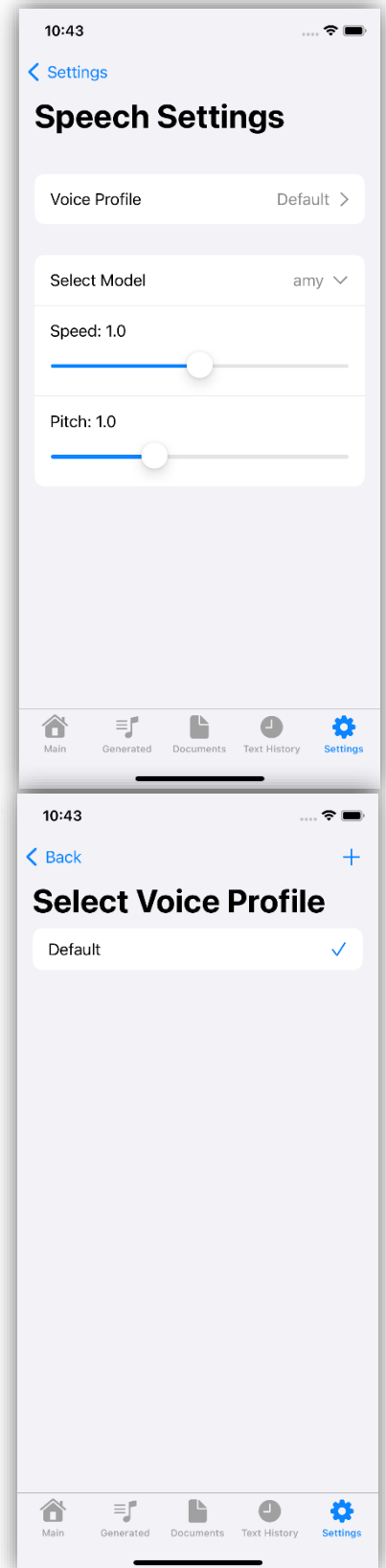
The Speech Settings section allows you to customize the voice and speaking parameters for text-to-speech (TTS) functionality. It includes options for selecting a voice profile, choosing an inference model, and adjusting playback characteristics such as speed and pitch.

To access the Speech Settings section:

1. Open the application.
2. Tap the gear icon located in the bottom-right corner of the home screen.
3. Select Speech Settings from the list of options.

3.5.3 Features in the Theme & Display Section

1. Voice Profile:
 - A Button to navigate to the Voice Profile selection and creation screen.
 - Click the “plus” sign on the top left to create and name a new voice profile.
 - Hold an existing voice profile to bring up the options to rename or delete voice profiles.
2. Select Model:
 - A button that brings up a dropdown menu listing available inference models.
3. Speed Slider:
 - A slider to adjust the speaking speed of the TTS engine.
 - Range: 0.5x (slower) to 1.5x (faster), with 1.0x as the default value.
4. Pitch Slider:
 - A button that brings up a dropdown menu listing available inference models.
 - Range: 0.5x (lower pitch) to 2.0x (higher pitch), with 1.0x as the default value.



3.5.3 Theme & Display

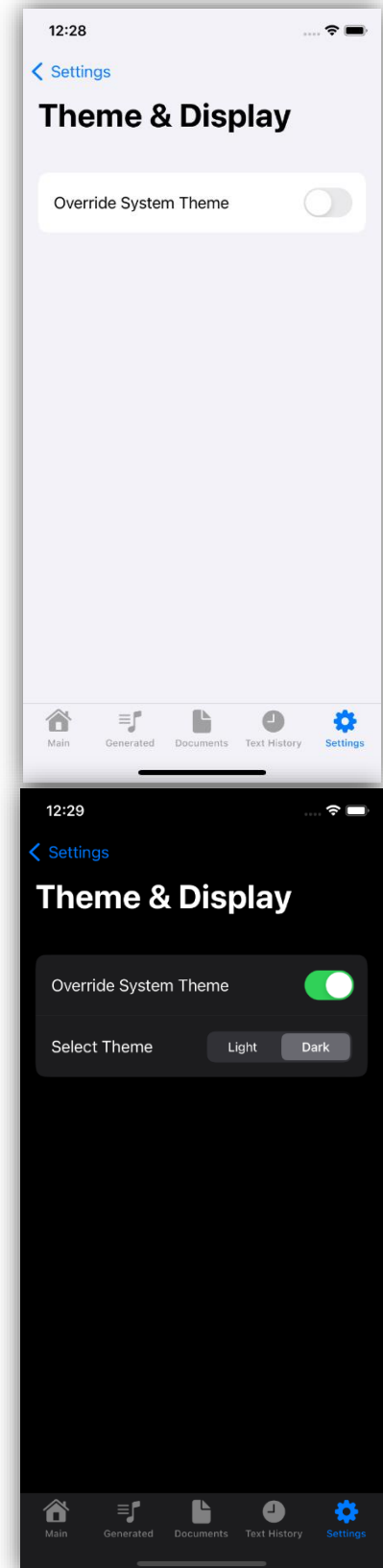
The Theme & Display section allows you to customize the app's visual appearance, providing options to align with your system settings or manually select between light and dark themes.

To access the Theme & Display section:

1. Open the application.
2. Tap the gear icon located in the bottom-right corner of the home screen.
3. Select Theme & Display from the list of options.

3.5.4 Features in the Theme & Display Section

1. Override System Theme:
 - A toggle switch to enable or disable manual theme selection.
2. Light/Dark Theme (Visible only when Override System Theme is enabled):
 - Allows manual switching between light and dark themes.
 - Light Theme: Bright background with dark text.
 - Dark Theme: Dark background with light text.



3.5.5 Security

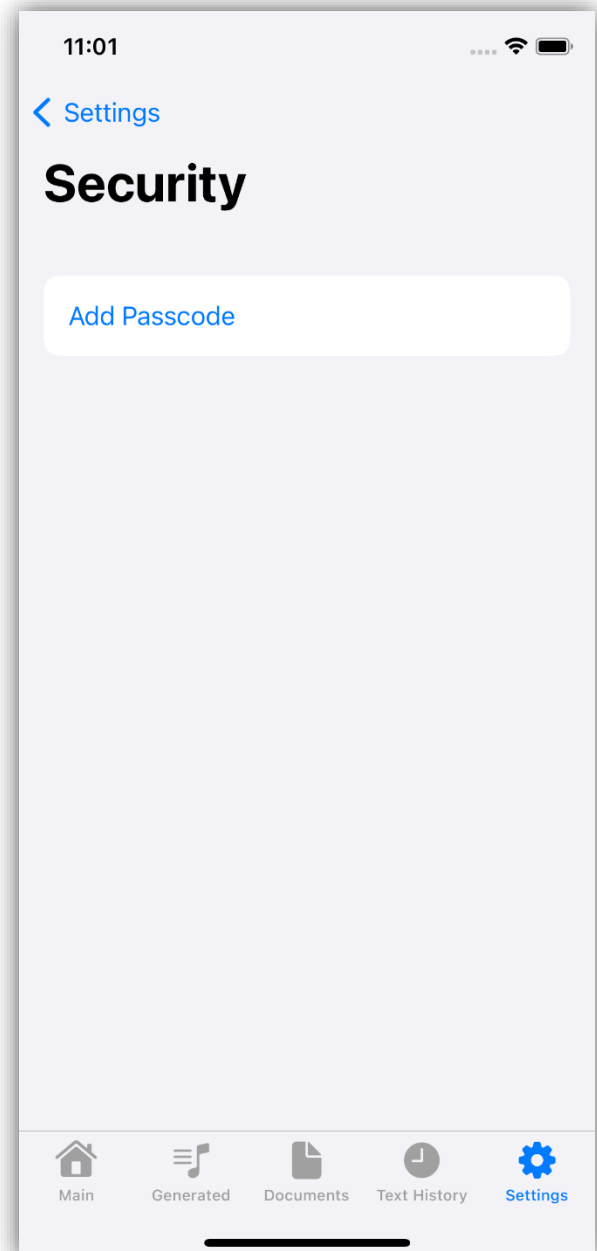
The Security section helps protect your app data by allowing you to manage a passcode. With options to add, change, or remove a passcode.

To access the Theme & Display section:

1. Open the application.
2. Tap the gear icon located in the bottom-right corner of the home screen.
3. Select Security from the list of options.

3.5.6 Features in the Security Section

1. Add Passcode:
 - Visible Only When No Passcode is Set.
 - A Button to navigate to the Add Passcode screen.
 - Enter a 4~6 digit passcode and confirm by entering it again.
 - Upon successful passcode creation, a Recovery Code will be presented.
 - Save the recovery code in a safe place. It is necessary if you forget your passcode, as the app does not store the passcode for security reasons.
2. Change Passcode:
 - Visible Only When a Passcode is Set.
 - A Button to navigate to the Add Passcode screen.
 - Enter your current passcode and new passcode twice to confirm the change.
3. Remove Passcode:
 - A Button to navigate to the Remove Passcode screen.
 - Enter your current passcode to confirm the action.





11:10

Cancel Save

Add Passcode

NEW PASSCODE

Enter New Passcode 

Confirm New Passcode 


Passcode must be 4 to 6 digits.

11:16


Cancel Save


Change Passcode

CURRENT PASSCODE

Enter Current Passcode 

NEW PASSCODE

Enter New Passcode 

Confirm New Passcode 


Passcode must be 4 to 6 digits.

11:16

Cancel

Remove Password

CONFIRM PASSCODE

Enter Current Passcode 

Remove Password

3.5.4 Advanced

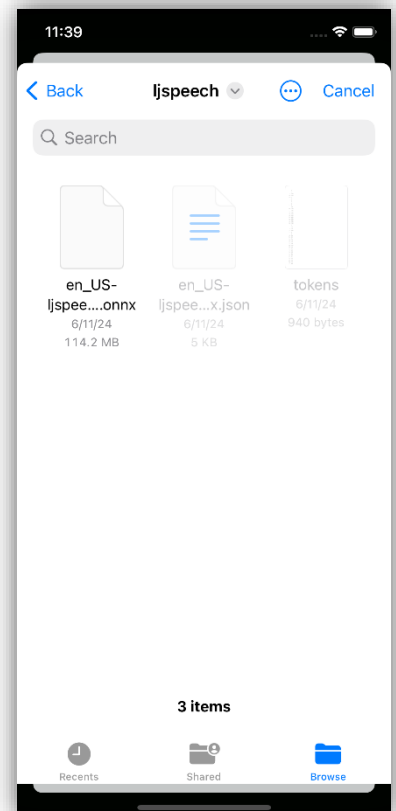
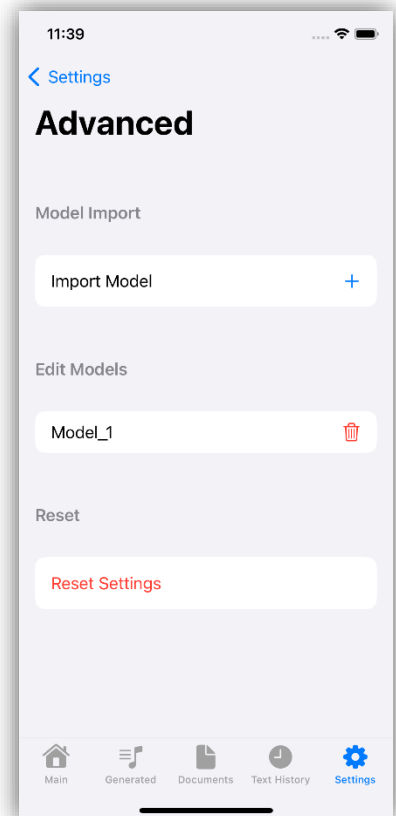
The Advanced section provides options for managing custom TTS models and resetting user preferences.

To access the Advanced section:

1. Open the application.
2. Tap the gear icon located in the bottom-right corner of the home screen.
3. Select Security from the list of options.

Features in the Advanced Section:

1. Import Model:
 - Allows for the import of a custom TTS model in the .onnx format into the app.
 - i. Tap the Import Model button to open the file system.
 - i. Navigate to and select desired .onnx file.
 - ii. Once imported, the model becomes available in the Speech Settings under the Select Model dropdown.
2. Edit Models:
 - Visible Only When There Is at Least One User-Imported Model.
 - Can be used to delete unwanted user-imported models.
3. Reset Settings:
 - Revert all user preferences to the default state.



3.5.5 Support

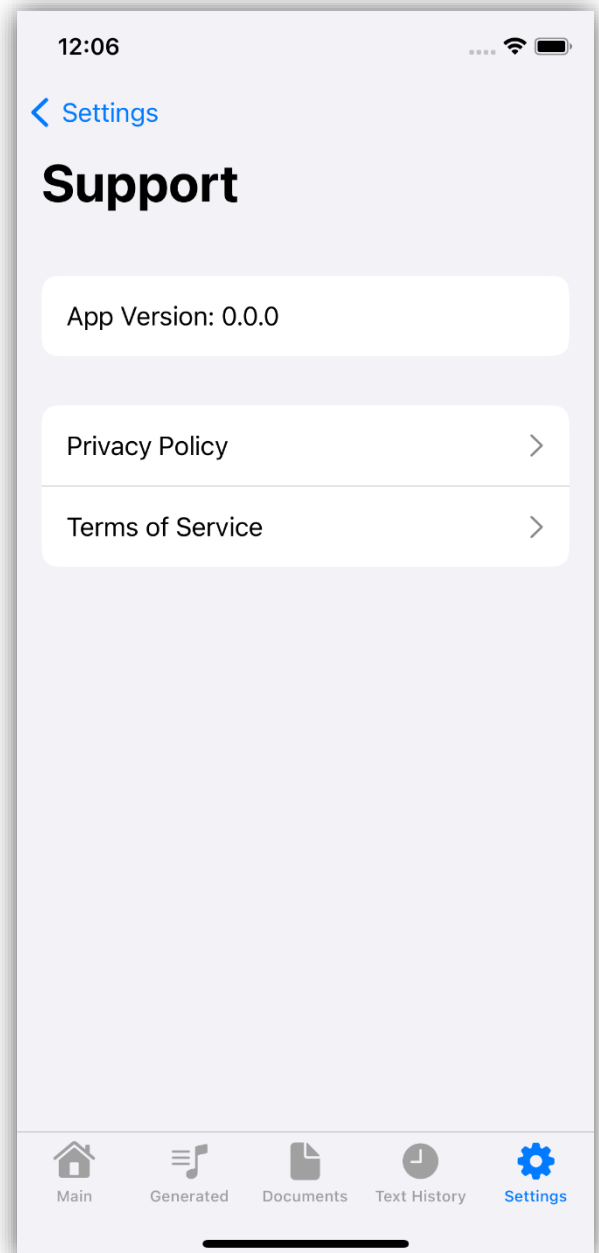
The Support section of EchoText provides helpful information about the application and documents like the Privacy Policy and Terms of Service. This section gives you quick access to legal information and application version details.

To access the Support section:

1. Open the application.
2. Tap the gear icon located in the bottom-right corner of the home screen.
3. Select Support from the list of options.

Features in the Support Section

1. App Version:
 - A text label at the top of the page displays the current version of the app.
 - Example: “App Version: 0.0.0”
 - This information can be useful when trouble shooting or checking for updates.
2. Privacy Policy:
 - Tapping this button will lead to the page with information about how the app handles user data.
5. Terms of Service:
 - Tapping this button will lead to the page with information outlining the rules and conditions for using the app.



4. Getting External Models

4.1 Getting VITS/Onnx model

To get a VITS (Variational Inference Text-to-Speech) model, you can follow these steps:

Pre-trained Models:

- Hugging Face: Users can find pre-trained VITS/ONNX models on platforms like Hugging Face. These models are ready to use and can be fine-tuned for specific needs (<https://huggingface.co/>).

Training a Model:

- Coqui TTS: Users can train their own VITS model using tools like Coqui TTS. Follow the official wiki which will guide you through the process of training a VITS model (https://docs.coqui.ai/en/latest/tutorial_for_nervous_beginners.html).

Community Resources:

- YouTube Tutorials: There are various YouTube tutorials, like the one on training a VITS model locally, which provide step-by-step instructions on how to prepare data, configure settings, and train the model.

4.2 Importing a VITS Model to ONNX using Python

To import a VITS model to ONNX, follow these steps:

Ensure you have the necessary libraries installed. You will need torch, onnx, and onnxruntime. You can install them using pip:

1. Ensure you have the necessary libraries installed. You will need torch, onnx, and onnxruntime. You can install them using pip:

```
pip install torch onnx onnxruntime
```

2. Load your pre-trained VITS model. If your model is saved as a .pth file, you can load it using PyTorch:

```
import torch
from your_model_file import VITSModel # Replace with your actual model file and class

model = VITSModel()
model.load_state_dict(torch.load('path_to_your_model.pth'))
model.eval()
```

3. Define the input and output tensors for the model. This is necessary for the ONNX export function:

```
dummy_input = torch.randn(1, 80, 100) # Adjust dimensions as per your model's requirements
```

4. Use the torch.onnx.export function to export the model to ONNX format:

```
torch.onnx.export(model, dummy_input, 'vits_model.onnx',
                  input_names=['input'], output_names=['output'],
                  dynamic_axes={'input': {0: 'batch_size'}, 'output': {0:
'batch_size'}})
```

5. Load and verify the ONNX model using onnxruntime to ensure it works correctly:

```
import onnx
import onnxruntime as ort

# Load the ONNX model
onnx_model = onnx.load('vits_model.onnx')
onnx.checker.check_model(onnx_model)

# Run a test inference
ort_session = ort.InferenceSession('vits_model.onnx')
outputs = ort_session.run(None, {'input': dummy_input.numpy()})
print(outputs)
```

This should help you get started with importing their VITS model to ONNX. If you need more detailed instructions or run into any issues refer to the official ONNX wiki.

5. Support

At EchoText, we're committed to ensuring that you have the best experience while using our application. If you encounter any issues, have questions, or want to provide feedback, our support team is here to help.

Contact Support - If you need further assistance, you can contact our support team:

- **Email:** hf3251@wayne.edu
Send a detailed description of your issue, including your device model, iOS version, and app version. Screenshots or error messages are helpful for faster resolution.