Personalized Automatic Speech Recognition

Application User Guide

Last revision: 20 August 2016

Table of Contents

Illustration Index

Intro

This is a user guide for the Personalized Automatic Speech Recognition desktop application which can be found here. In this guide you can find information regarding the different scenes of the application and how to use them.

General

The application consists of five different scenes namely:

- **Initial Scene**
- **E-mail List Scene**
- **LDA Scene**
- **Record Scene**
- **Dictate Scene**

In the flow chart below you can see the possible transitions between these five scenes:

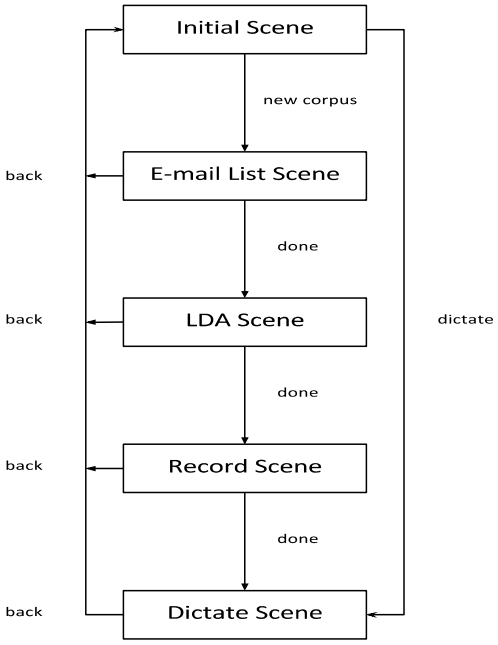

Illustration 1: Scene transition flow chart

Console

Besides the main application window where the five main scenes appear, there is a secondary window, the console window. On this window you can see messages sent from the application, mainly when something has gone wrong or a requested action could not be performed. While it is safe to put this window aside, check on it from time to time so you don't miss any message.

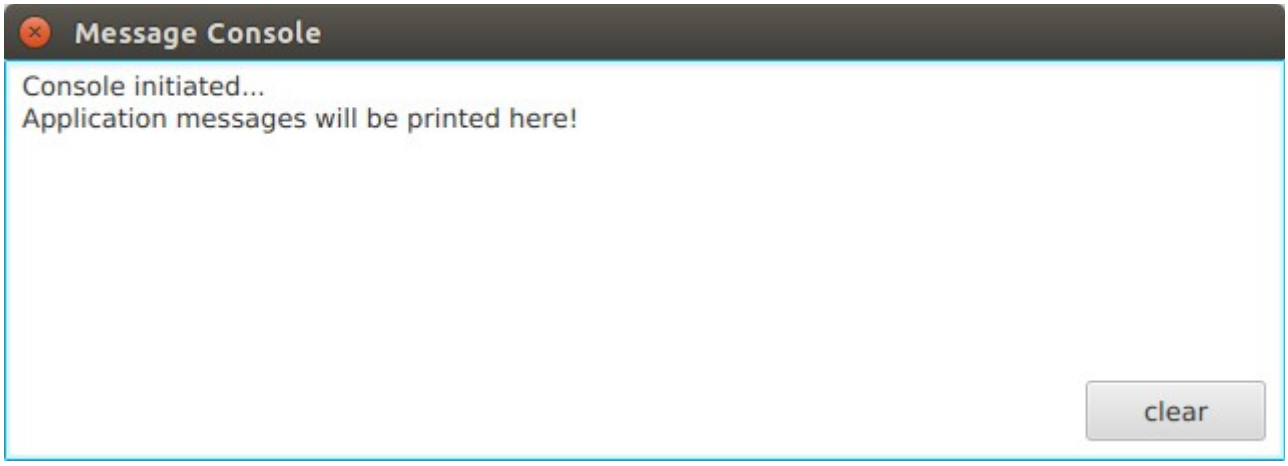

Illustration 2: Console window at the moment the application is started

Initial Scene

This is the first scene that will be displayed by the application. It consists of two parts, the left and the right part.

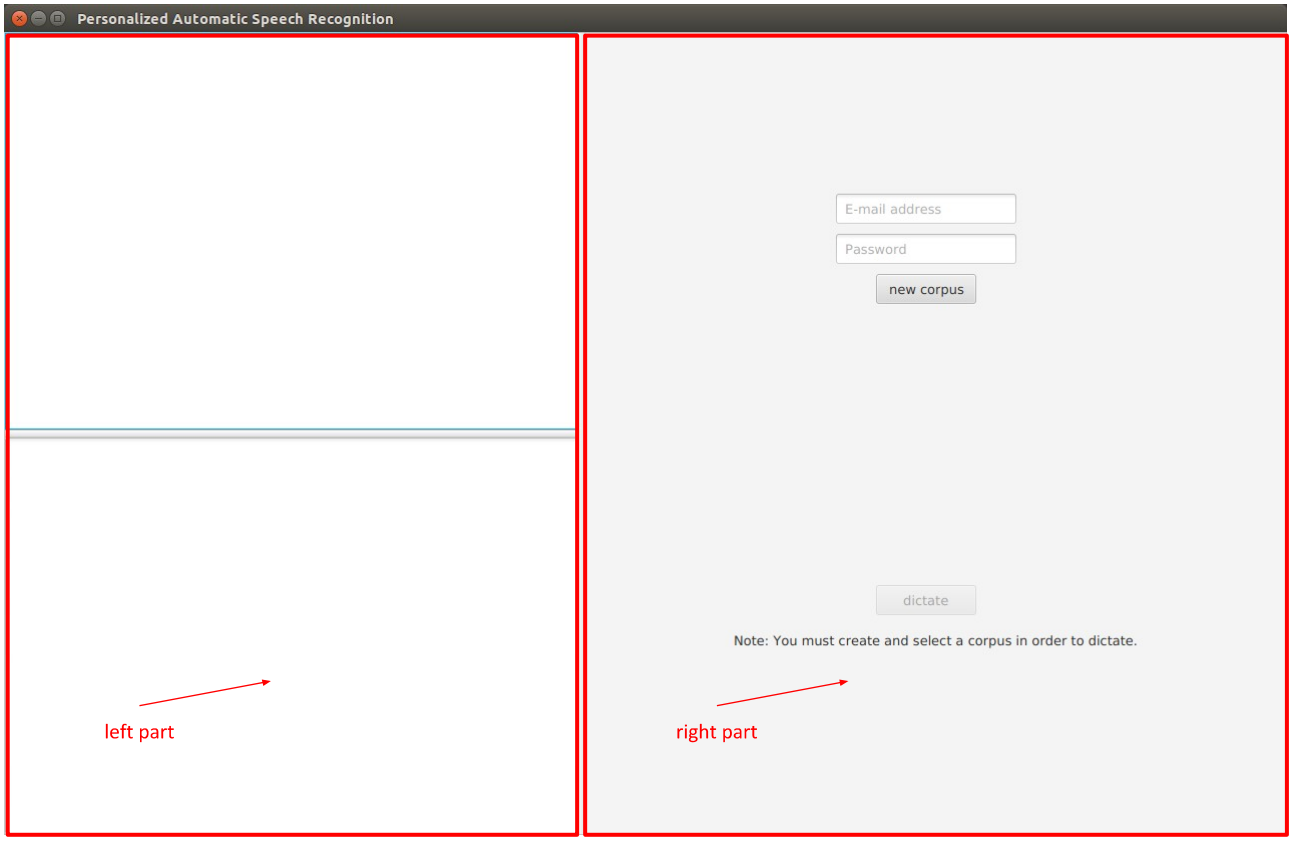

Illustration 3: Initial Scene

On the left part, you will see a list with the corpora you have created and saved during previous runs of the application. Concretely, on the upper left part you will see a list with the names of the available corpora while on the lower left part you will see the actual text of a corpus once you have selected it on the upper part's list. The first time you run the application, there are no corpora

saved so the list will be empty.

On the lower right part you can see the dictate button. This button will take you to the Dictate Scene once you have selected a corpus.

On the upper right part you can see the e-mail address and password text fields along with the new corpus button. This is where you want to go the first time you run the application. Entering your e-mail address and password and clicking the new corpus button will take you to the E-mail List Scene.

The following notes should be made here:

1. At the moment, only gmail addresses are supported.
2. Your password is not saved anywhere so you will have to provide it each time you want to connect to your e-mail address.
3. The connection to the gmail server is done through an SSL connection.
4. Your e-mails are being parsed by the application and saved inside the database directory of the application for the sole purpose of speech recognition.
5. You e-mails are only read and not modified.

Entering a wrong e-mail address – password combination will result in a message being posted on the console.

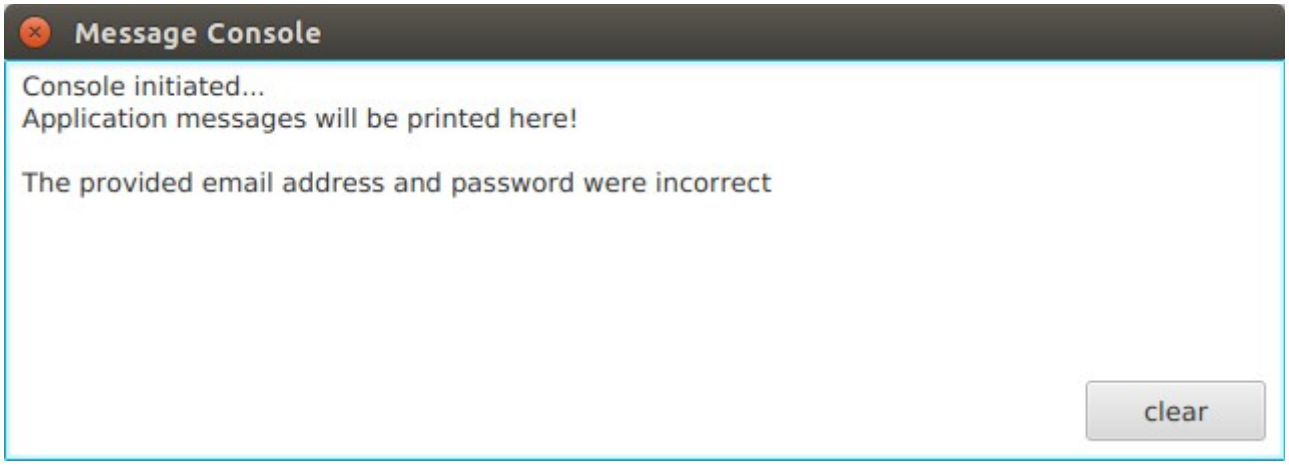

Illustration 4: Console window with an error message posted

E-mail List Scene

This scene will appear after you have successfully entered your e-mail address and password on the Initial Scene. This scene consists of two parts, the left and the right part.

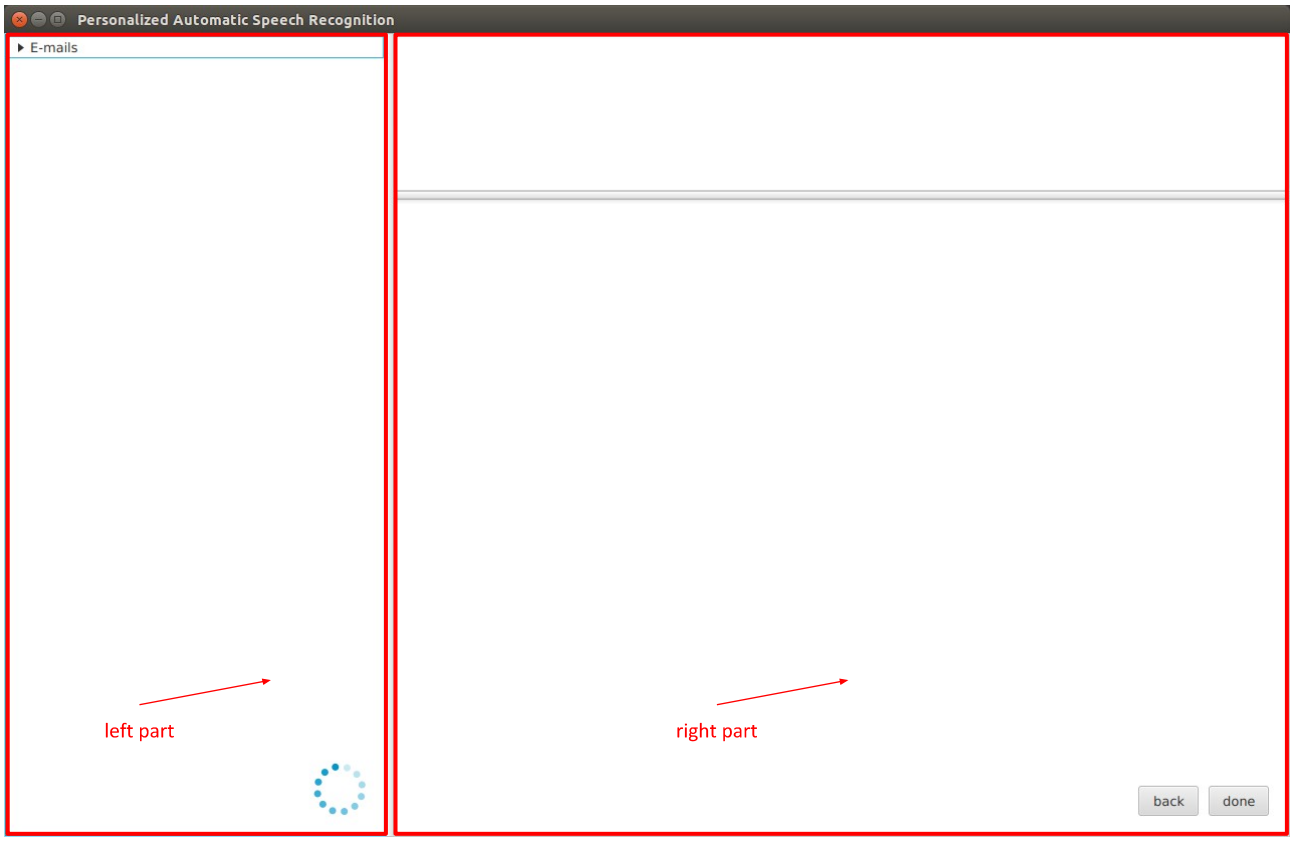

Illustration 5: E-mail List Scene

On the left part you will see your e-mails as they are being fetched from the remote server. Notice the spinner on the lower right corner of the left part. This spinner indicates that the e-mails are being fetched. It will disappear once all your e-mails have been fetched.

The e-mails are presented in a tree-like list. To expand an e-mail folder, click on the little arrow on the left of the folder.

Each time you select an e-mail you will see its contents on the right part. Concretely, you will see its body on the lower right part and information regarding the subject, the senders, the receivers, etc... on the upper right part.

You can select multiple e-mails by holding down the control button of your keyboard. If you select a folder, then all the e-mails inside the folder will be selected (selecting both a folder and an e-mail inside the folder will lead to the e-mail being selected twice). Once you have selected all the e-mails you want to use to create a new corpus, click the done button. This will take you to the LDA Scene.

Clicking the back button will take you back to the Initial Scene.

LDA Scene

This scene will appear after you have chosen one or more e-mails on the E-mail List Scene and clicked the done button. This scene consists of two parts, the left and the right part.

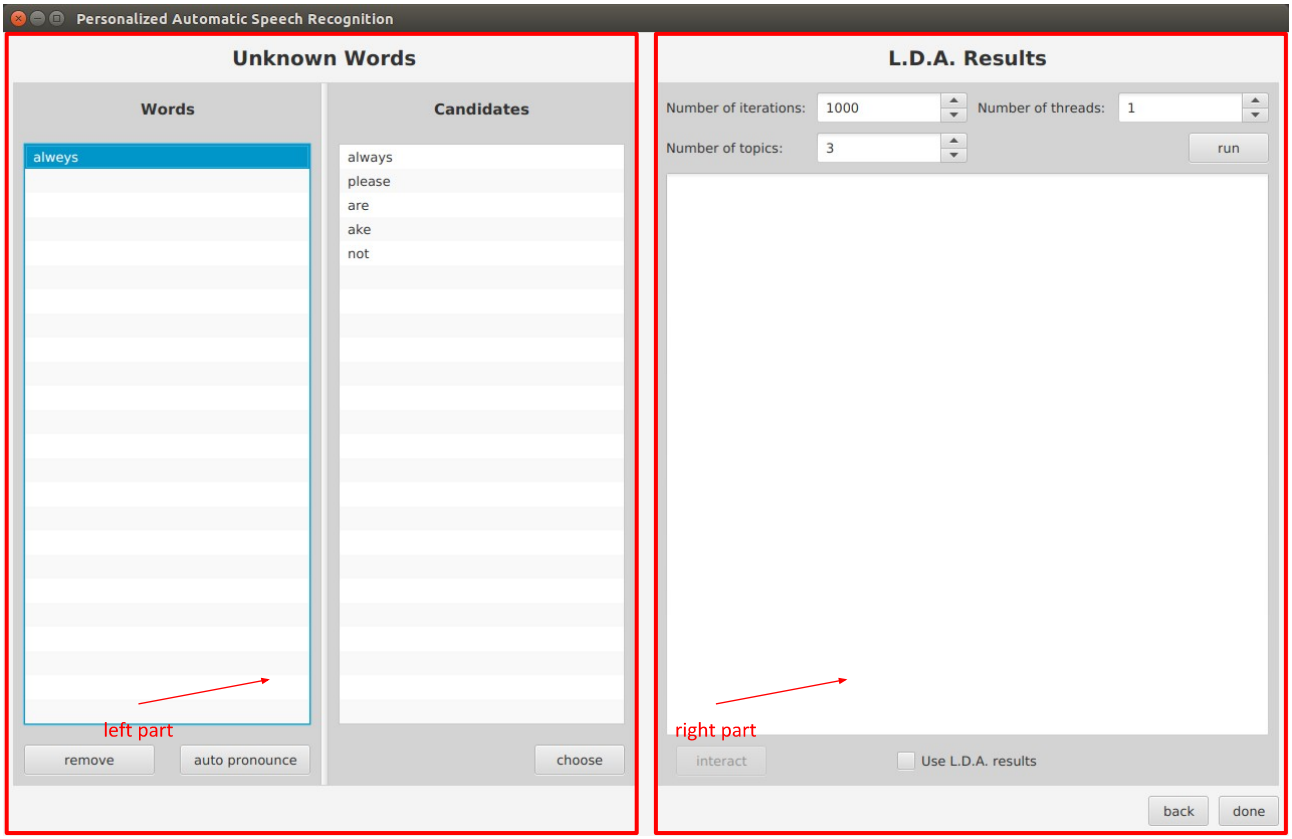

Illustration 6: LDA Scene

The left part will help you deal with the unknown words that have been found in your e-mails. The application uses an English dictionary with more than 100.000 words so there might be words that are not included into the dictionary. A misspelled word will also be detected as unknown so you will have the chance to correct it. For each unknown word, a list of possible substitutes will be provided. You can also choose to auto pronounce an unknown word (the aforementioned dictionary is actually a pronunciation dictionary that maps words to pronunciation sequences, for example, the -> DH AH). Finally, you can choose to simply remove an unknown word.

Note: Removing an unknown word will only remove it from the local copy of your e-mails. The e-mails saved on the remote server remain intact and are never changed by the application.

The right part of the scene will apply the LDA algorithm on the selected e-mails. This algorithm will try to detect different topics on your e-mails and cluster them in topic-specific groups. This means that, if you choose an e-mail that contains a discussion about science and another e-mail that contains a discussion about religion, these two e-mails will be assigned to a different group.

You can parameterized the LDA algorithm by setting the number of iterations (the more iterations you add the more accurate results you will get but it will take more time to execute), the number of topics to search for and the number of thread to use (use more than one only if you have chosen ten or more e-mails since the parallelization overhead makes the execution slower for less e-mails).

After the LDA algorithm is finished, the results will be presented to you inside the text area on the

right part of the scene. You can click the interact button to see the topic each e-mail has been assigned to. You can even change the topic of each e-mail or assign an e-mail on more than one topics.

After you have finished, you should check the “use LDA result” check box if you want to use the aforementioned topics. Concretely, checking the box will result in the creation of many corpora, one for each topic while leaving the box unchecked will result in the creation of only one corpus will all the chosen e-mails in it.

Click the done button to indicate that you are ready to move to the Record Scene.

Clicking the back button will take you to the Initial Scene.

Record Scene

On this scene you can record samples of your voice which will be used for adapting the default acoustic model of CMU Sphinx to your voice and surrounding environment in order to improve the automatic speech recognition results. Your voice will be recorded while you are dictating a specified sentence. This scene consists of two parts.

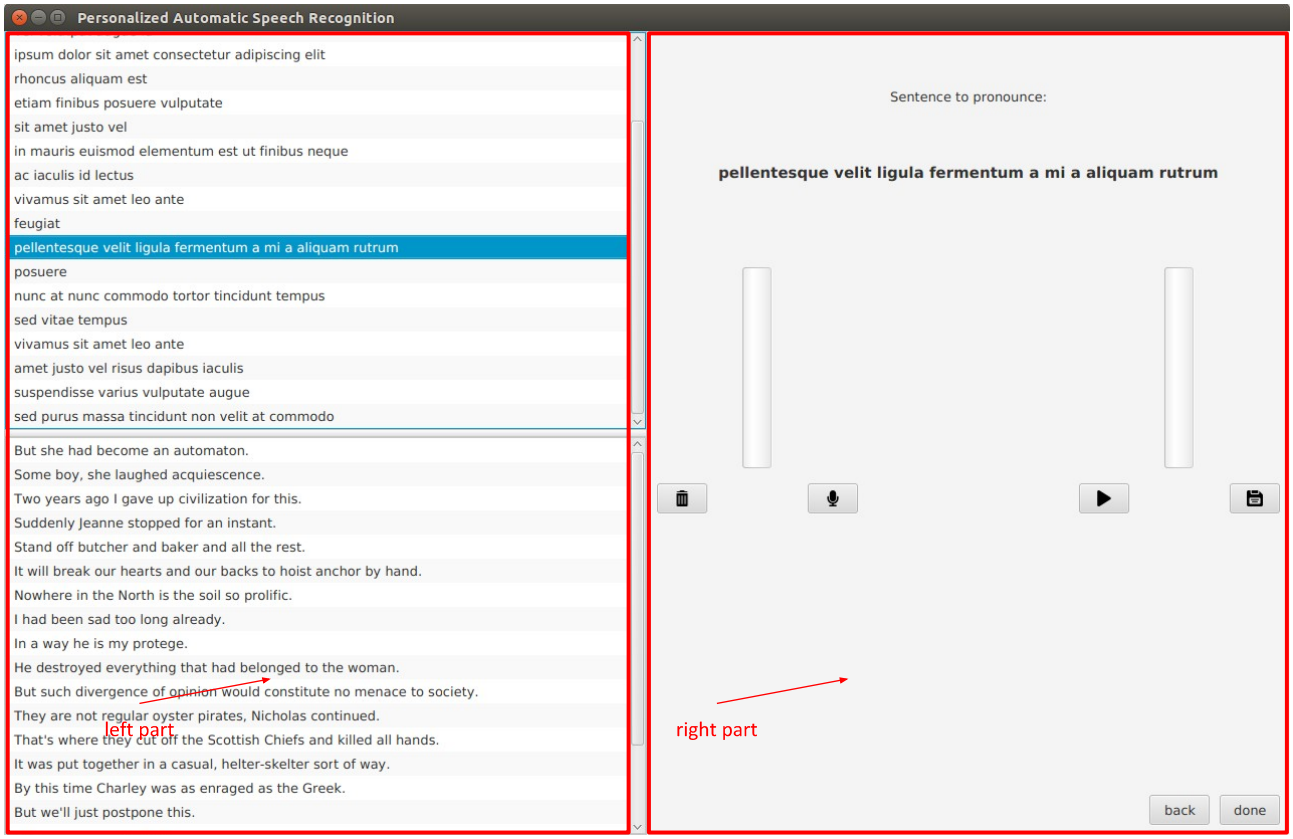

Illustration 7: Record Scene

On the left part of the scene you can see two lists of sentences. The upper list contains sentences chosen from your corpus while the lower list contains sentences from the CMU ARCTIC data.

In order to record a voice sample, you should first choose a sentence from either of the two lists, and then click the record button on the right part of the scene. Click the record button again to stop the recording. In order to listen your sample, you can choose the play button on the right part of the scene. To save the sample click the save button or click the erase button to delete it.

In order to get actual improvement on the recognition accuracy, you should provide several samples. You will need at least five minutes of recorded speech which can be roughly translated to one hundred samples. You don't have to provide all the samples at once. The samples you provide are not corpus specific which means that, if you create five corpora and provide twenty samples for each, you will have provided the one hundred required samples.

Clicking the back button will take you to the Initial Scene while clicking the done button will take you to the Dictate Scene.

Dictate Scene

On this scene you can dictate a sentence and the application will return the recognized transcript to you. The scene consists of two parts, the left and the right part.

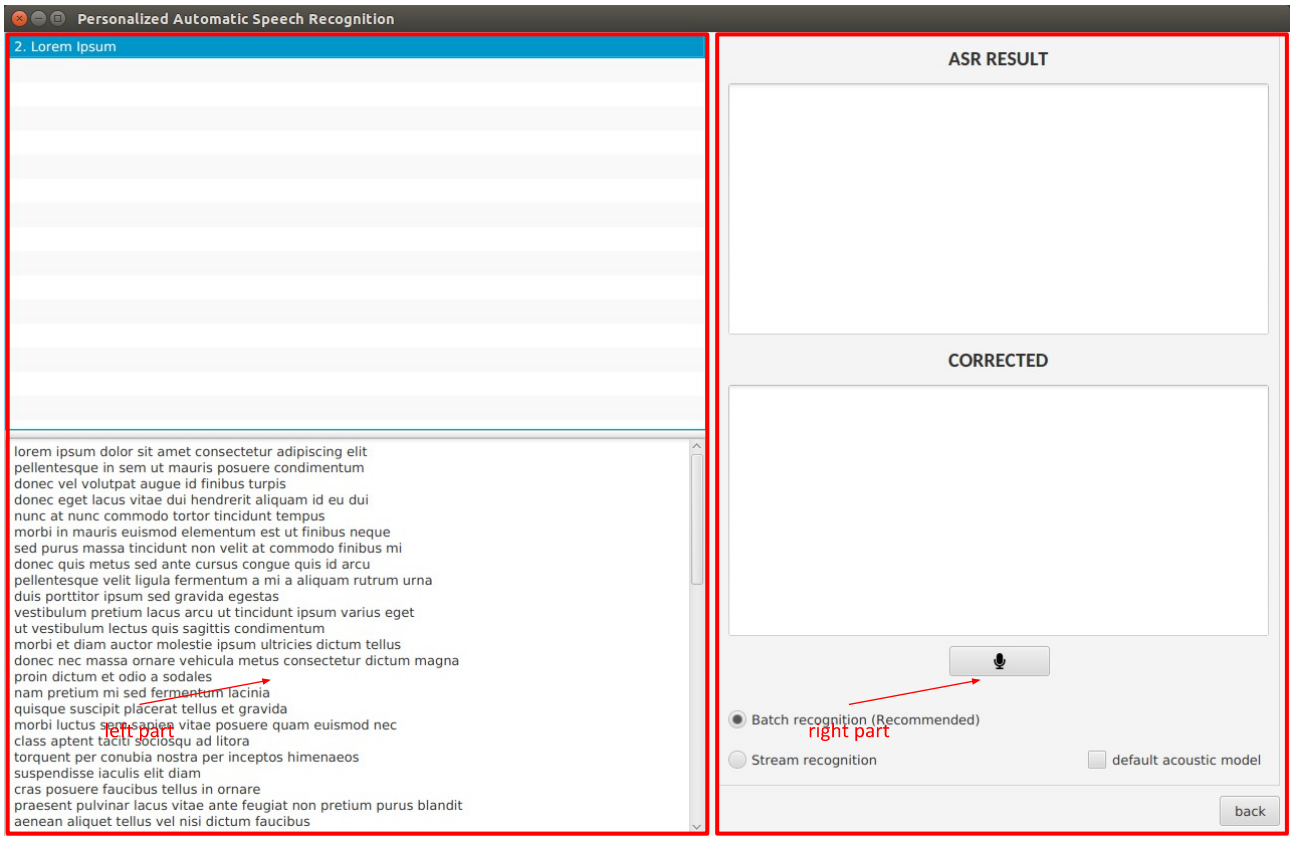

Illustration 8: Dictate Scene

On the left part of the scene you will see a list of the corpora you have created just like the left part of the Initial Scene. Choosing a corpus will configure the application to recognize sentences from it. It might take a few seconds for the configuration to be done.

On the right part of the scene, you can see two text areas. On the upper text area you will see the recognized transcript as it was returned by the CMU Sphinx recognition engine. On the lower text area you will see the recognized transcript as it was returned by the correction engine of the application. The text contained on the lower text area, should be considered the final result of the application.

On the right part of the scene, you can also see a check box with the label “default acoustic model”. Checking this check box will configure the application to recognize using the acoustic model as it is provided by CMU Sphinx, ignoring all the samples you provided on the Record Scene.

There are also two radio buttons. Choosing the “Batch recognition” will configure the application to start the recognition after you have finished your dictation. This means that you will have to click the dictate button, dictate your sentence and then click again the dictate button in order for the recognition to begin. Choosing the “Stream recognition” will configure the application to recognize your speech in real time. This means that you will see the recognized transcript while you dictate. Tests has shown that batch recognition yields better results and thus is the recommended way to go.

In order to dictate, you should click the dictate button. Depending on the recognition style you have chosen, you will see the recognized text appear in real time (Stream recognition) or after you have clicked the dictate button again (Batch recognition).

Clicking the back button will take you to the Initial Scene.