# ASR Suite for Dummies

## Overview

This document is a software manual, describing the tools used to yoke the simulated hearing aid to a model of the patients impaired hearing (the dummy), and how the dummy can be subsequently connected to an automatic speech recogniser. The idea is that if an automatic speech recogniser is trained using feature vectors derived from a model of normal hearing, then the speech recogniser will perform optimally when tested with feature vectors derived from the same normal-hearing dummy. If the recogniser is tested using feature vectors derived from an impaired dummy, then one would expect the recognition score to drop significantly. The representation of the acoustic stimulus can be modified by a hearing aid algorithm before being presented to the impaired dummy. The potential of the hearing aid signal processing to alleviate some of the reduction in speech intelligibility caused by the impairment may then be assessed by comparing the aided and unaided recognition scores. This software framework should eventually enable the tailoring of hearing aid parameters to individuals in their absence.

The software suite exists in a “userPrograms” folder within the main Matlab model of the Auditory Periphery (MAP) folder. This tutorial document assumes that the reader is already familiar and comfortable with using MAP. The full documentation for MAP can be found at:

<http://www.essex.ac.uk/psychology/department/research/hearing_models.html>

## Files included in the suite

The evaluation of the hearing aid is done with the assistance of 3 main classes.

### cEssexAid.m

This is the file that contains the class definition of the Essex Aid wrapper. The Essex Aid is a novel hearing aid algorithm developed at the University of Essex. Once the user is familiar with the software framework, it should be relatively straightforward to swap out the hearing aid algorithm so that any hearing aid algorithm can be tested.

### cJob.m

This is the class definition that is the real workhorse of the operation. It provides many utility functions to assist the user in creating feature sets for use in training and testing the speech recogniser. This class also contains code for scheduling, so that many sounds can be processed in parallel on one or many machines should you have the resources available to do so.

### cHMM.m

This is a wrapper class for the Hidden Markov Tookkit (HTK) <http://htk.eng.cam.ac.uk/>. This class contains helper functions that provide the recogniser with the necessary information for training a HMM and scoring the results produced.

\*\*\*

Other functions within the parent directory include

### worker.m

This is an autonomous little function that takes a path to a folder containing a list of jobs as an input argument. A job is defined here as a list of wav files that need to each be converted into a feature vector. It searches the directory for jobs to do and works until all of the jobs are complete.

### Exp\_Tutorial\_X.m

Files prefixed with “Exp\_” are the files that users will edit most frequently. They are at the top of the function call stack and initiate all of the tasks required to run a recognition experiment.

### MAPwrap.m

This is a very simple wrapper for MAP, making it easy to call from the classes within the suite. MAP also uses a lot of global variables, so the wrapper also provides a barrier to prevent potential conflicts between variable names.

\*\*\*

There are also two folders in the parent directory

### /def

This contains the definition files required by HTK, such as grammar rules, dictionaries and hmm prototypes. Full descriptions of these files are beyond the scope of this document, but detailed information can be found here <http://htk.eng.cam.ac.uk/>

### /ASRfiles

This is a folder containing some additional utility functions used by the main classes.

\*\*\*

## Tutorial

One of the best ways to learn is by doing. The tutorial is split into two sections that represent a typical workflow. The first part of the tutorial shows the reader how to make a new HMM and then test it. The second part shows the reader how to use an existing HMM to test new features. The new data might have come from a different dummy and/or hearing aid processing.

Before any work with the recogniser can be accomplished, the speech material and recogniser software must be in place.

### Install HTK

HTK needs to be compiled for your platform and added to the path. The method for doing this will be different under different operating systems. See the following link for information on how to add programs to the path under Windows.

<http://lmgtfy.com/?q=windows+add+to+path>

Once the HTK binaries have been successfully added to the path, the individual tools should be available from Matlab. This can be tested by issuing the following command (the >> should not be typed):

>> !HVite -V

This command should then output some version information about HTK into the command window. If the command fails to return version information, then logging off and then back on again should solve the problem.

### Get appropriate speech material

The software provided is designed to work with the AURORA 2.0 TI digits corpus available here:

<http://www.elda.org/article52.html>

The clean training data should be in wav format and placed into one directory. The clean, digit-triplet test sound files should placed in a separate directory.

It is also possible to make a custom corpus so long as the following rules and file naming conventions are adhered to.

1. Speech material should be recorded in (or converted to) wav file format. Any sampling rate can be used, as the Matlab scripts will resample the speech files appropriately. The speech should be recorded in single channel format.
2. Training data should be recorded as strings of digits between 1 and approximately 7 digits per file. Test data files should contain 3 digits. Digits should include a fairly even mixture of “oh”, “one”, “two”, “three”, “four”, “five”, “six”, “eight”, “nine”. The bisyllabic digits “seven” and “zero” should not be used.
3. Speech files should be named like the following example, “FAC\_8O4A.wav”. The first character in the string refers to the gender of the talker. The next two characters are a unique identifier for the specific person doing the talking. These two characters should be followed by an underscore. The next characters are the string of numbers that are uttered in the sound file. These numbers are terminated with a capital “A” and the wav extension. IMPORTANT NOTE: If the file contains the utterance “oh”, the alphabetic character “O” should be used rather than the numerical character “0”.
4. The files in the training and testing corpora should be unique.

### Get appropriate noise material

Any noise material of suitable duration (substantially longer than the longest sound file containing speech) can but used. The tutorials here use the “factory1” noise sample from the freely available NOISEX database.

<http://spib.rice.edu/spib/select_noise.html>

### Show the cJob class where to find the sound files

The main job class (cJob.m) needs to know the location of the speech material. The speech material can be stored at any location on the user’s computer, but the following piece of code needs to be updated accordingly. If different computers on different platforms are used then the paths can be set accordingly. For example, I use a windows machine in the office, a mac at home, and a linux machine to run large jobs. All of these systems store the corpora in different locations.

if isunix

if ismac

lWAVpath = '~/ASR/reducedAURORA/TrainingData-Clean/';

rWAVpath = '~/ASR/reducedAURORA/TripletTestData/';

obj.noiseFolder = '~/ASR/noises';

else

lWAVpath = '/scratch/nnn/corpora/AURORA digits (wav)/TrainingData-Clean/';

rWAVpath = '/scratch/nnn/corpora/AURORA digits (wav)/TripletTestData/';

obj.noiseFolder = '/scratch/nrclark/corpora/noises';

end

else

lWAVpath = 'C:\corpora\AURORA digits (wav)\TrainingData-Clean';

rWAVpath = 'C:\corpora\AURORA digits (wav)\TripletTestData';

obj.noiseFolder = 'C:\corpora\noises';

end

### Setting an output folder

The speech recognition experiments involve generating a large number of files that need to be stored somewhere. It is possible to change the output folder on an experiment by experiment basis, but the user may wish to have a top level directory in which all ASR data is stored. To do this, find and amend the following code segment.

else

if isunix

if ismac

obj.opFolder = '~/ASR/exps/\_foo';

else

obj.opFolder = '/scratch/nrclark/exps/\_foo';

end

else

obj.opFolder = 'D:\exps\\_foo';

end

end

### Training and testing a recogniser

If everything has been set correctly, it should now be possible to run Exp\_Tutorial\_1 without generating errors. This function can be used as a template to run all kinds of recognition experiments. The following text breaks down each part of the function, describing what happens in each block of code.

The first line of the file declares the function and states that it takes one input argument, isMasterNode.

function Exp\_Tutorial\_1(isMasterNode)

Speech recognition experiments are very processor intensive and so the software suite has been carefully designed to run in parallel across many instances of matlab. This allows results to be generated in a fraction of the time of a serial process if enough computing power is available. The simple scheduling software was written in house, and so it does not require any special matlab licenses for clustering. The variable isMasterNode is a Boolean variable that lets the current matlab instance know if it is the master node. The master node is the most important node, responsible for generating all of the job information and interfacing with HTK. These responsibilities will be discussed in detail at the appropriate point in the code.