We will go through the background of both of the challenges in detail on the first morning of the hackathon so don’t worry if any of the terms used here are unfamiliar. This is just an overview so you know what format of data to expect.

**Software needed:**

Please download Anaconda3 version 5.2.0 from here: <https://repo.anaconda.com/archive/>

Anaconda contains most of the packages that we will use for our example models but you will also need to install tensorflow and sklearn packages by opening the anaconda terminal and typing:

conda install tensorflow

pip install sklearn

### If you have any questions or problems please email [Jessica.monaghan@mq.edu.au](mailto:Jessica.monaghan@mq.edu.au) or [Jorge.mejia@nal.gov.au](mailto:Jorge.mejia@nal.gov.au)

**Challenge 1**

How well can speech understanding be predicted from electroencephalogram (EEG) data?

There are two EEG datasets, either one or both of these datasets can be used in the challenge. The datasets will be released on the morning of the first day after the presentations. The aim is to use feature vectors derived from the EEG data to predict speech understanding in terms of percentage of words or morphemes (parts of words) correctly identified. A subset of the subjects will be withheld as a test set. The EEG data and features from the test set will be released on the morning of the third day and you will provide us with your predicted speech scores for evaluation of your models.

**Dataset 1** comes from a study by Miles *et al.* (2017) <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5536372/>.

The stimuli in this experiment were simple sentences which have been degraded by channel vocoding (<https://en.wikipedia.org/wiki/Vocoder>) and the addition of noise. There are four conditions with different difficulties (6 channel vocoding, low noise condition; 6 channel vocoding high noise condition; 16 channel vocoding low noise condition; 16 channel vocoding, high noise condition). There are 55 trials for each condition and 23 participants overall. While EEG was recorded the participants listened to each sentence and once it ended tried to repeat the sentence as accurately as possible. The performance measure to be predicted is percentage of words correctly identified in each sentence. For each trial, the power present in each frequency band (alpha, beta, gamma, delta, theta) is available along with a baseline value used for normalisation. Additionally, the filtered EEG data for each trial and channel are provided.

Data fields:

'eegwholetrial' – filtered EEG data (0.5 Hz – 45 Hz) – 6 seconds duration, 27 channels

'fs' – sampling rate (= 500 Hz)

'eeg\_Alpha\_Active' – 27 dimensions (channels) per trial

'eeg\_Alpha\_Baseline'– 27 dimensions (channels) per trial

'eeg\_Beta\_Active'– 27 dimensions (channels) per trial

'eeg\_Beta\_Baseline'– 27 dimensions (channels) per trial

'eeg\_Delta\_Active'– 27 dimensions (channels) per trial

'eeg\_Delta\_Baseline'– 27 dimensions (channels) per trial

'eeg\_Theta\_Active'– 27 dimensions (channels) per trial

'eeg\_Theta\_Baseline'– 27 dimensions (channels) per trial

'eeg\_Gamma\_Active'– 27 dimensions (channels) per trial

'eeg\_Gamma\_Baseline'– 27 dimensions (channels) per trial

'SpeechRecall' – number of words correctly identified per trial

'ToSpeechItems' – total number of words per trial

'condition' – 1 to 4

'siglen' – sentence durations (s)

'sigonset' – sentence onset relative to trial (s)

'sex' – 1 or 0

'age'

*Miles K., McMahon C., Boisvert I., Ibrahim R., de Lissa P., Graham P., Lyxell B. (2017) Objective assessment of listening effort: Coregistration of pupillometry and EEG. Trends in Hearing 21: 2331216517706396.*

**Dataset 2**

The data presented in the second dataset is from three different studies, one of which was published by Mejia *et al*. (2017) <http://www.hearingreview.com/2017/01/listening-effort-speech-intelligibility-narrow-directionality/#>. In all studies, the target stimuli selected were simple sentences presented in speech babble background noise.  Participants were asked to recall what they heard. The performance measure to be predicted is percentage of morphemes (parts of words) correctly identified in each sentence. The signal to noise ratios used were 0, -3.5 and -5 dB SNR for the different studies. In all studies, an EEG electrode head cap was used to record EEG signals, while participants performed the recall task. The number of trials per participant varied from 42 and up to 64. For each trial, the power present in each frequency band (alpha, beta, gamma, delta, theta) is available along with a baseline value used for normalisation. Additionally, the raw EEG data for each trial and channel are provided. In total, data from 56 participants are made available for this challenge.

Data fields:

'eegwholetrial' – unfiltered EEG data – 6 seconds duration, 6 channels

'fs' – sampling rate (= 500 Hz)

'eeg\_Alpha\_Active' – 6 dimensions (channels) per trial

'eeg\_Alpha\_Baseline'– 6 dimensions (channels) per trial

'eeg\_Beta\_Active'– 6 dimensions (channels) per trial

'eeg\_Beta\_Baseline'– 6 dimensions (channels) per trial

'eeg\_Delta\_Active'– 6 dimensions (channels) per trial

'eeg\_Delta\_Baseline'– 6 dimensions (channels) per trial

'eeg\_Theta\_Active'– 6 dimensions (channels) per trial

'eeg\_Theta\_Baseline'– 6 dimensions (channels) per trial

'eeg\_Gamma\_Active'– 6 dimensions (channels) per trial

'eeg\_Gamma\_Baseline'– 6 dimensions (channels) per trial

'SpeechRecall' – number of words correctly identified per trial

'ToSpeechItems' – total number of words per trial

'sex' – 1 or 0

'age'

*Mejia J, Carter L, Dillon H, Littman V. Listening Effort, Speech Intelligibility, and Narrow Directionality. Hearing Review. 2017; 24(1):22*

**Challenge 2**

Can we determine what factors are causing hearing loss in young people, using a dataset of 1400 respondents that includes lifestyle surveys, physiological measurements and hearing tests?

The clinical data structure shown here was extracted from a large-scale hearing health study, involving participants 11 – 35 years of age. The study was funded by the Australian Commonwealth Government. Researchers collected data from 1400 participants, and the data collected include audiometric, physiological measures and lifestyle surveys. Below is a summary of the data structure used in this challenge.

% Clinical ID record of participant

MyIHeardata.SubjID

% Conventional audiometric data record

MyIHeardata.General.Gender % Male 1, Female 0

MyIHeardata.General.Age

MyIHeardata.PTA.Left % Freq: 500, 1000, 2000, 3000, 4000, 6000, 8000

MyIHeardata.PTA.Right % Freq: 500, 1000, 2000, 3000, 4000, 6000, 8000

% Otoscopy:  is an examination that involves looking into the ear with an

% instrument called an otoscope(or auriscope)

% Data in binary form = Ok 0, Issues 1

MyIHeardata.Otoscopy.Wax\_Left

MyIHeardata.Otoscopy.Wax\_Right

MyIHeardata.Otoscopy.Discharge.Left

MyIHeardata.Otoscopy.Discharge.Right

MyIHeardata.Otoscopy.Perforation.Left

MyIHeardata.Otoscopy.Perforation.Right

MyIHeardata.Otoscopy.stenosis.Left

MyIHeardata.Otoscopy.stenosis.Right

MyIHeardata.Otoscopy.Atresia.Left

MyIHeardata.Otoscopy.Atresia.Right

MyIHeardata.Otoscopy.MiddleEar\_dysfunction.Left

MyIHeardata.Otoscopyy.MiddleEar\_dysfunction.Right

MyIHeardata.Otoscopy.Infection.Left

MyIHeardata.Otoscopy.Infection.Right

MyIHeardata.Otoscopy.HistoryOf\_ME\_Disfunction.Left

MyIHeardata.Otoscopy.HistoryOf\_ME\_Disfunction.Right

% General Health related issues

MyIHeardata.Health.HeadTrauma

MyIHeardata.Health.RiskFactors

MyIHeardata.Health.AtBirth

MyIHeardata.Health.FamilyHistory

MyIHeardata.Health.DrugsExposure

MyIHeardata.Health.ChemicalsExposure

% Distortion product otoacoustic emissions (DPOAEs)

% reflects outer hair cell integrity and cochlear function. When used

% appropriately in the audiology clinic, they are an effective diagnostic

% tool and can detect hearing loss with accuracy

% instrument called an otoscope(or auriscope)

% continuous variables

MyIHeardata.OAS.P1.DPNF.Left % Noise floors Band 1, Band 8

MyIHeardata.OAS.P1.Amplitude.Left % Amplitude floors Band 1, Band 8

MyIHeardata.OAS.P1.DPNF.Right % Noise floors Band 1, Band 8

MyIHeardata.OAS.P1.Amplitude.Right %Amplitude floors Band 1, Band 8

% P stands for protocol 2

MyIHeardata.OAS.P2.DPNF.Left

MyIHeardata.OAS.P2.Amplitude.Left

MyIHeardata.OAS.P2.DPNF.Right

MyIHeardata.OAS.P2.Amplitude.Right

% SNR : Signal to noise ratio

% Protocol 1

MyIHeardata.OAS.P1.SNR.Left

MyIHeardata.OAS.P1.SNR.Right

% Protocol 2

MyIHeardata.OAS.P2.SNR.Left

MyIHeardata.OAS.P2.SNR.Right

%% Transient Evoked Otoacoustic Emissions (TEOAE)

% The test result is represented by a curve that simply shows whether

% optoacoustic emission is present or not.  if the otoacoustic emissions are

% present it means adequate functioning of the cochlea.

% continuous variables

MyIHeardata.TEL.SNR.Left

MyIHeardata.TEL.Amp.Left

MyIHeardata.TER.SNR.Right

MyIHeardata.TER.Amp.Right

MyIHeardata.TEL.NF.Left % Equipment noise floor

MyIHeardata.TEL.NF.Right

%% Adults data: dosimeter data pascal-squared-hours

% Personal sound systems: continuous variables

MyIHeardata.Adults.Dosimeter.personal\_stereo.moreThan60dBSPL.Left

MyIHeardata.Adults.Dosimeter.personal\_stereo.moreThan60dBSPL.Right

MyIHeardata.Adults.Dosimeter.personal\_stereo.moreThan65dBSPL.Left

MyIHeardata.Adults.Dosimeter.personal\_stereo.moreThan65dBSPL.Right

MyIHeardata.Adults.Dosimeter.personal\_stereo.moreThan70dBSPL.Left

MyIHeardata.Adults.Dosimeter.personal\_stereo.moreThan70dBSPL.Right

MyIHeardata.Adults.Dosimeter.personal\_stereo.moreThan75dBSPL.Left

MyIHeardata.Adults.Dosimeter.personal\_stereo.moreThan75dBSPL.Right

MyIHeardata.Adults.Dosimeter.personal\_stereo.moreThan80dBSPL.Left

MyIHeardata.Adults.Dosimeter.personal\_stereo.moreThan80dBSPL.Right

MyIHeardata.Adults.Dosimeter.personal\_stereo.moreThan85dBSPL.Left

MyIHeardata.Adults.Dosimeter.personal\_stereo.moreThan85dBSPL.Right

MyIHeardata.Adults.Dosimeter.personal\_stereo.moreThan90dBSPL.Left

MyIHeardata.Adults.Dosimeter.personal\_stereo.moreThan90dBSPL.Right

MyIHeardata.Adults.Dosimeter.personal\_stereo.moreThan95dBSPL.Left

MyIHeardata.Adults.Dosimeter.personal\_stereo.moreThan95dBSPL.Right

MyIHeardata.Adults.Dosimeter.personal\_stereo.moreThan100dBSPL.Left

% Other activities

% continuous variables

MyIHeardata.Adults.Dosimeter.OtherLeisure.moreThan60dBSPL

MyIHeardata.Adults.Dosimeter.OtherLeisure.moreThan65dBSPL

MyIHeardata.Adults.Dosimeter.OtherLeisure.moreThan70dBSPL

MyIHeardata.Adults.Dosimeter.OtherLeisure.moreThan75dBSPL

MyIHeardata.Adults.Dosimeter.OtherLeisure.moreThan80dBSPL

MyIHeardata.Adults.Dosimeter.OtherLeisure.moreThan85dBSPL

MyIHeardata.Adults.Dosimeter.OtherLeisure.moreThan90dBSPL

MyIHeardata.Adults.Dosimeter.OtherLeisure.moreThan95dBSPL

MyIHeardata.Adults.Dosimeter.OtherLeisure.moreThan60dBA\_SPL

MyIHeardata.Adults.Dosimeter.OtherLeisure.moreThan65dBA\_SPL

MyIHeardata.Adults.Dosimeter.OtherLeisure.moreThan70dBA\_SPL

MyIHeardata.Adults.Dosimeter.OtherLeisure.moreThan75dBA\_SPL

MyIHeardata.Adults.Dosimeter.OtherLeisure.moreThan80dBA\_SPL

MyIHeardata.Adults.Dosimeter.OtherLeisure.moreThan85dBA\_SPL

MyIHeardata.Adults.Dosimeter.OtherLeisure.moreThan90dBA\_SPL

MyIHeardata.Adults.Dosimeter.OtherLeisure.moreThan100dBA\_SPL

%% Questionnaire: Ordinal scores, 1 Lowest (No) to 5 Highest

MyIHeardata.Adults.Questionnaire.Tinitus.Left

MyIHeardata.Adults.Questionnaire.Tinitus.Right

MyIHeardata.Adults.Questionnaire.AboriginalTorresStraitIslander

MyIHeardata.Adults.Questionnaire.EducationalLevel

MyIHeardata.Adults.Questionnaire.Trouble.inQuiet

MyIHeardata.Adults.Questionnaire.Trouble.inNoise

MyIHeardata.Adults.Questionnaire.Trouble.inLectures

MyIHeardata.Adults.Questionnaire.Noticed\_Hearing\_Change

MyIHeardata.Adults.Questionnaire.Concerned\_Hearing\_Getting\_Worse

MyIHeardata.Adults.Questionnaire.Believes.Hearing\_Could\_Change

MyIHeardata.Adults.Questionnaire.Believes.Loud\_Noisy\_Leisure\_Increases\_Risk

MyIHeardata.Adults.Questionnaire.Believes.Loud\_Noisy\_Headphone\_Increases\_Risk

MyIHeardata.Adults.Questionnaire.Believes.Loud\_NightClubs\_Increases\_Risk

MyIHeardata.Adults.Questionnaire.Awareness.Loud\_Noise\_Risk

MyIHeardata.Adults.Questionnaire.Awareness.Loud\_Noise\_Problems

MyIHeardata.Adults.Questionnaire.Behaviour.LoudSpeaker\_Music\_Usage

MyIHeardata.Adults.Questionnaire.Behaviour.LoudSpeaker\_Music\_Loudness

MyIHeardata.Adults.Questionnaire.Behaviour.Avoids\_Loud\_Sounds

%% School age: dosimeter data pascal-squared-hours

% continuous variables

MyIHeardata.SchoolAge.Dosimeter.School.moreThan60dBA\_SPL

MyIHeardata.SchoolAge.Dosimeter.School.moreThan65dBA\_SPL

MyIHeardata.SchoolAge.Dosimeter.School.moreThan70dBA\_SPL

MyIHeardata.SchoolAge.Dosimeter.School.moreThan75dBA\_SPL

MyIHeardata.SchoolAge.Dosimeter.School.moreThan80dBA\_SPL

MyIHeardata.SchoolAge.Dosimeter.School.moreThan85dBA\_SPL

MyIHeardata.SchoolAge.Dosimeter.School.moreThan95dBA\_SPL

MyIHeardata.SchoolAge.Dosimeter.School.moreThan100dBA\_SPL

% 1 Aboriginal, 2 Torres Strait Islander, NaN neither

MyIHeardata.SchoolAge.Questionnaire.AboriginalTorresStraitIslander

%% Questionnaires: Ordinal scores, 1 Lowest (No) to 5 Highest

MyIHeardata.SchoolAge.Questionnaire.Tinitus.Left

MyIHeardata.SchoolAge.Questionnaire.Tinitus.Right

MyIHeardata.SchoolAge.Questionnaire.Trouble.inQuiet

MyIHeardata.SchoolAge.Questionnaire.Trouble.inNoise

MyIHeardata.SchoolAge.Questionnaire.Trouble.inClass

MyIHeardata.SchoolAge.Questionnaire.Noticed\_Hearing\_Change

MyIHeardata.SchoolAge.Questionnaire.Concerned\_Hearing\_Getting\_Worse

MyIHeardata.SchoolAge.Questionnaire.Believes.Hearing\_Could\_Change

MyIHeardata.SchoolAge.Questionnaire.Believes.Loud\_Noisy\_Leisure\_Increases\_Risk

MyIHeardata.SchoolAge.Questionnaire.Believes.Loud\_Noisy\_Headphone\_Increases\_Risk

MyIHeardata.SchoolAge.Questionnaire.Awareness.Loud\_Noise\_Risk

MyIHeardata.SchoolAge.Questionnaire.Awareness.Loud\_Noise\_Problems

MyIHeardata.SchoolAge.Questionnaire.Behaviour.LoudSpeaker\_Music\_Usage

MyIHeardata.SchoolAge.Questionnaire.Behaviour.LoudSpeaker\_Music\_Loudness

MyIHeardata.SchoolAge.Questionnaire.Behaviour.Avoids\_Loud\_Sounds

%% Personal sound systems: dosimeter data pascal-squared-hours

% continuous variables

MyIHeardata.SchoolAge.Dosimeter.personal\_stereo.moreThan60dBA\_SPL.Left

MyIHeardata.SchoolAge.Dosimeter.personal\_stereo.moreThan60dBA\_SPL.Right

MyIHeardata.SchoolAge.Dosimeter.personal\_stereo.moreThan65dBA\_SPL.Left

MyIHeardata.SchoolAge.Dosimeter.personal\_stereo.moreThan65dBA\_SPL.Right

MyIHeardata.SchoolAge.Dosimeter.personal\_stereo.moreThan70dBA\_SPL.Left

MyIHeardata.SchoolAge.Dosimeter.personal\_stereo.moreThan70dBA\_SPL.Right

MyIHeardata.SchoolAge.Dosimeter.personal\_stereo.moreThan75dBA\_SPL.Left

MyIHeardata.SchoolAge.Dosimeter.personal\_stereo.moreThan75dBA\_SPL.Right

MyIHeardata.SchoolAge.Dosimeter.personal\_stereo.moreThan80dBA\_SPL.Left

MyIHeardata.SchoolAge.Dosimeter.personal\_stereo.moreThan80dBA\_SPL.Right

MyIHeardata.SchoolAge.Dosimeter.personal\_stereo.moreThan85dBA\_SPL.Left

MyIHeardata.SchoolAge.Dosimeter.personal\_stereo.moreThan85dBA\_SPL.Right

MyIHeardata.SchoolAge.Dosimeter.personal\_stereo.moreThan90dBA\_SPL.Left

MyIHeardata.SchoolAge.Dosimeter.personal\_stereo.moreThan90dBA\_SPL.Right

MyIHeardata.SchoolAge.Dosimeter.personal\_stereo.moreThan95dBA\_SPL.Left

MyIHeardata.SchoolAge.Dosimeter.personal\_stereo.moreThan95dBA\_SPL.Right

MyIHeardata.SchoolAge.Dosimeter.personal\_stereo.moreThan100dBA\_SPL.Left

MyIHeardata.SchoolAge.Dosimeter.personal\_stereo.moreThan100dBA\_SPL.Right

**PUBLISHED PAPERS**

*Carter, L., Williams, W., Black, D., & Bundy, A. (2014). The leisure-noise dilemma: hearing loss or hearsay? What does the literature tell us? Ear and Hearing 35(5): 491505. doi:10.1097/01.aud.0000451498.92871.20*

*Carter, L., Black, D., Bundy, A., & Williams, W. (2016a). An estimation of the wholeof-life noise exposure of adolescent and young adult Australians with hearing impairment. Journal of the American Academy of Audiology 27: 1-14. doi: 10.3766/jaaa.15100*

*Carter, L., Black, D., Bundy, A., & Williams, W. (2016b). Parent perceptions of children's leisure and the risk of damaging noise exposure. Deafness & Education International 18(2): 87-102. doi: 10.1080/14643154.2015.1136478*