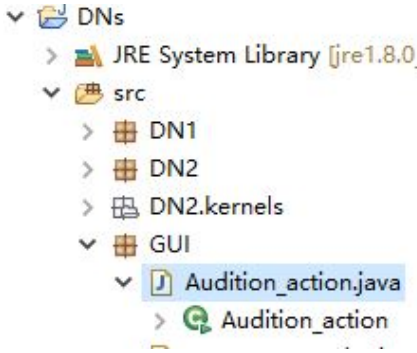
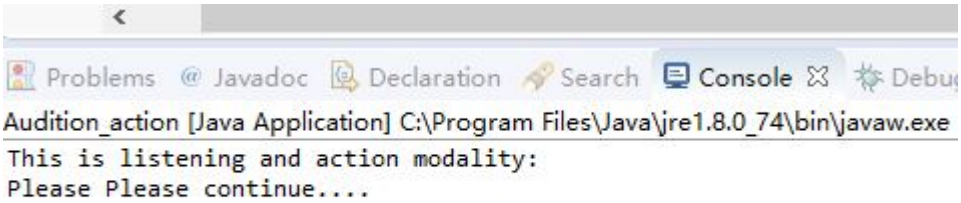
Documentation of phoneme experiment

1. Running the phoneme experiment

Right click the Audition_action.java.



And press any key to continue.



Run as Java application. The output file “actions.txt” is used to reconstructed the synthesized sounds with the Matlab code “5_reconstruct.m”.

The raw data processing and corresponding PCA states (generated by CCI PCA) are the Matlab codes (see “DNs/Data_processing/READ_ME.txt”).

2. Settings

2.1 Directory: DNs/Audition_Data/DN2 /action/Input

This directory contains all the speech data the network will use.

There are 6 segments, the first segment has a length of 8 and other segments have a length of 5614 samples with input retention. These segments are:

(a) pretraining-> All motors are supervised, for training volume.

(b) training -> All motors are supervised.

(c) resubstitution -> Only the first two samples are supervised.

(d) disjoint 1 ~disjoint 3 -> New sequence (different samples from (b)) where only the first two samples are supervised.

This directory contains the “settings_audition.txt” file. Here we specify the network's size for each Area, number of segments, and length of each segment.

2.1 Data Settings

The input data set used in this problem contains 2 parts. For each frame, the first part is the input feature matrix and the second part is the corresponding volume vectors.

Each input frame represents 20ms waveform segment (with 10ms overlap between neighbor frames). The feature matrix is 11*10, means the original waveform is filtered by a series of sine functions.

Each volume vector in the frame indicates the volume of the frame. The volume vector is 10*8, which is extended by a 1*4 vector.

For the 1*10 vector, each bit represents one moving direction.

If the 1^{st} bit is one, the frame is silence; If the 2^{nd} bit is one, the volume of the frame is low; If the 3^{rd} bit is one, the volume of the frame is middle; If the 10^{th} bit is one, the volume of the frame is high.

For the motor area, there are one concept zone and 60 real value sections. The concept zone is the type of phonemes, Each real value section represents a real value in the range of -1 to 1. All 60 real value sections represent a 1*60 vector.

We use 1*21 vector to represent 20 types of phonemes and silence concept for the type concept. We set 1*800 vector for the second concept zone. The 1*60 vectors generated by CCI PCA are fed to the 60 real value sections

2.2 Training and tests process

In the pre-training stage, we train DN-2 to learn the local features. We use randomly recordings with different volume to train DN-2, DN-2 will learn to distinguish silence and sound.

In the training stage (sequence 2), we train DN-2 to learn the phoneme sequences. DN-2 leans to synthesize and recognize different phonemes.

When tests, we use one re-substitution sequence and 3 disjoint sequences to test DN-2.