
This document is also available in PDF¹ format.

The data directory <http://www.dsp.sun.ac.za/pr813/data/> contains various data sets that will be used in this course. Here is a short description of each set:

- The **simvowel set**² contains the files `klasxa.txt` to `klasxu.txt` as training data. The corresponding testing data is in `testxa.txt` to `testxu.txt`. The features are four-dimensional. Do not for one minute believe that this data really is a simulation of anything; it is purely contrived data useful for illustrating some concepts.
- The **timit set**³ contains speech data useful for speaker and phone recognition. Speakers are arranged according to “dialect regions” `dr1` through `dr8`. For *speaker recognition* purposes we will use the `dr1` data in the `train` subdirectory. Each final subdirectory in `dr1` indicates a specific speaker or class. The `.cep` files contain 16-dimensional cepstral features calculated from 16ms frames of speech. See the `read.me` file in the `timit` directory for file formats. The supplied `loadjdp.m` function can read the data into Matlab. Use the `sx*.cep` files for training the system, and then test it on the `si*.cep` and `sa*.cep` files.
The data set is also useful for *phoneme recognition*. In this case each `.cep` file is viewed as a collection of phonemes, as indicated by the corresponding `.phn` file. The supplied `timitphon.m`⁴ creates a training and test set from all examples of a specific phoneme. Good phonemes to highlight the advantage of HMMs over GMMs include the diphthongs and stops (see `phoncode.doc` for more information). We will use the diphthongs `ey`, `aw`, `ay`, `oy` and `ow` as five classes. The examples found in the `sx*.cep` files serve as training set, while the test examples are retrieved from the `si*.cep` and `sa*.cep` files.
- The **timit_phone_sim set**⁵ is a version simulating the effects of a telephone channel on timit. This is done by low-pass filtering it to 4KHz and then adding noise for a resultant signal-to-noise ratio (SNR) of about 25dB.
- The **timit_hf set**⁶ is an awefullized version of `dr1` timit. Transmitted via the HF radio channel from Pretoria to Stellenbosch - the effect is about the same that your luggage will experience if you dragged your suitcase on a chain behind your car over the same distance.
- The **faces set**⁷ is a database of images of human faces from Tom Mitchell’s website⁸. Each final subdirectory indicates a particular person. Each person is photographed from four different angles (straight, left, right, up), with four expressions (happy, sad, neutral, angry), and with/without sunglasses. We will concentrate on the set of “straight” faces, with each person considered a class. For *face recognition* experiments, train your algorithms on the happy and sad

¹<http://www.dsp.sun.ac.za/pr813/datasets/pr813.datasets.pdf>

²<http://www.dsp.sun.ac.za/pr813/data/simvowel.tar.gz>

³<http://www.dsp.sun.ac.za/pr813/data/timit.tar.gz>

⁴<http://www.dsp.sun.ac.za/pr813/data/timitphon.m>

⁵http://www.dsp.sun.ac.za/pr813/data/timit_phone_sim.tar.gz

⁶http://www.dsp.sun.ac.za/pr813/data/timit_hf.tar.gz

⁷<http://www.dsp.sun.ac.za/pr813/data/faces.tar.gz>

⁸<http://www-2.cs.cmu.edu/afs/cs.cmu.edu/user/mitchell/ftp/faces.html>

expressions, and then test it on the angry and neutral ones. Another useful experiment which allows more training data per class is *pose recognition*. In this case there are four classes (straight, left, right, up), with the first 15 people selected as training set and the remaining five as test set.