

Consonant Challenge: baseline recogniser and training scripts

Ning Ma, Martin Cooke and Odette Scharenborg
{n.ma, m.cooke}@dcs.shef.ac.uk, o.scharenborg@let.ru.nl

March 2008

1 Introduction

This package contains an HMM-based speech recogniser and training scripts for use in the Consonant Challenge, Interspeech 2008. This distribution is based on HTK V3.2 and employs monophone models and MFCC features for modelling consonants.

While extensive testing has been performed on these unix shell scripts, we would appreciate feedback on any problems that you encounter. Please report problems to Ning Ma (n.ma@dcs.shef.ac.uk) in the first instance.

2 Contents of the distribution

The **chome** directory contains 6 subdirectories:

etc/	Config files for the recogniser and training scripts
flists/	Feature file lists for various data sets
labels/	Master label files (MLFs)
models/	Monophone HMM definitions
results/	Recognition results
scripts/	Scripts for feature extraction, HMM training and the baseline recogniser

3 Feature extraction

Script file: **scripts/do_hcopy.sh**; script variables needing to be changed: **\$HTKPATH**, **\$DATAROOT**.

The script performs feature extraction and takes a single argument: *setname* (e.g. *train*, or *testset1*). It assumes that HTK binaries can be found in **\$HTKPATH/** and wav files (.wav) can be found in **\$DATAROOT/wav/setname**. The encoded features (MFCC in this distribution) will be saved in the directory **\$DATAROOT/mfcc/setname**.

4 Training scripts

Script file: `scripts/do_htk_train.sh`; script variables needing to be changed: `$HTKPATH`, `$FTRROOT`.

The script performs training for monophone HMMs up to 24 mixtures using HTK. Feature files (.mfcc) are assumed to be found in the directory `$FTRROOT/`. The script employs *HCompV* to initialise monophone models and *HERest* to perform EM training. The final HMMs are saved in the directory `models/`.

5 Baseline recogniser

Script file: `scripts/do_htk_recog.sh`; script variables needing to be changed: `$HTKPATH`, `$FTRROOT`.

The baseline recogniser takes a *setname* (e.g. *testset1*) as its argument and assumes that feature files (.mfcc) for the dataset can be found in `$FTRROOT/`. The recognition results will be saved in the directory `results/`.

Standard MFCC features (13-dimensional) are employed together with delta and acceleration coefficients (see `etc/config_hcopy_mfcc` for details) resulting in a 39-dimensional feature vector.

30 monophone models were employed: 24 consonants and two models for each of the 3 vowels used in this challenge – one to model the initial vowel context and the other one to model the final vowel context of the VCV. Each of the monophones consists of 3 emitting states with a 24-Gaussian mixture output distribution. No silence model and short pause model are employed in this distribution. A set of trained HMMs (24-Gaussian mixture) is included in this package and can be found at `models/HMMs_monophone_24mix_mfcc`.

HTK commands *HVite* and *HResults* are employed to perform recognition and calculate accuracies. If you employ the set of models included in this package, an accuracy of **88.54%** should be achieved for *testset1*. The .mlf and .hresults files for *testset1* using the models are supplied in the directory `results/mfcc/`.