READ ME Mocha-Dansa V 1.0

Veena Singampalli

January 14, 2009

Contents

1 Overview

This document describes the changes made to the label files in version 1.2 of MOCHA-TIMIT database for speakers msak0 and fsew0. The structure and the contents of the folders, the format of label files and changes made are explained here.

2 Directory structure

The label files are stored in <*msak0_Dansa_1-0*> for male speaker and <*fsew0_Dansa_1-0*> for female speaker. The information of the exact changes made to the label files is specified in the form of excel spreadsheets, *msak0_Dansa_1-0.xls* and *fsew0_Dansa_1-0.xls* for male and female speakers respectively.

3 File format

Every label file is in the HTK format and is stored with an extension ".lb". The start times and end times are in 10^{-7} units. For example, msak0_001.lb is the label file for 1st sentence uttered by speaker msak0 beginning:

```
0         3000000   sil
3000000   5100000   breath
5100000   5800000   sil
...
```

4 Changes made to the label files

The information on changes made to the label files of both speakers is stored in two spreadsheets. Each excel file has two sheets named **Corrections** and **Duration changes**. **Corrections** lists the labelling errors and the modifications made, **Duration changes** lists the changes made to the total duration of the label files for consistency with the EMA data.

4.1 Correction of labelling errors

This section presents the two types of labelling errors encountered in version 1.2 of MOCHA-TIMIT, (a) the whole sentence had wrong labels with entries for word(s) missing from the transcriptions, (b) existing label files were adjusted taking into consideration errors from failed forced alignment, elision and feature spread (for alveolar consonants [t], [d] and [n])

The sheet **Corrections** specifies the file name, start times (in 10ms) ST, end times (in 10ms) ET, affected phones, the transcription for that word before and after the changes, the orthography, observational comments and the action taken. Where applicable, measurement errors (already mentioned in MOCHA-TIMIT V.1.2) are additionally specified. The observational comments were guided by inspection of spectrogram of the acoustic file along with the EMA measurements. This study was carried out only on the alveolar stops in the database.

4.1.1 Wrongly labelled files

Sentences 173, 317, 332, 340, 352, 354, 357, 369 had entries missing for a significant portion of the actual utterances. New labels were generated for each sentence based on the phonetic transcriptions of the words and aligned to acoustic input manually.

4.1.2 Files that needed adjustments

Labelling errors were also observed due to failed forced alignment, elision and the feature spread. In case of failed forced alignment, adjustments were made to the phone boundaries manually by identification of the correct starting and end times with the help of spectrograms. Entries for elided phones were deleted from the transcriptions and the labels for the neighbouring phones were adjusted based on the spectrogram and the EMA readings. Realisation of nasal sound [n] was found to be affected by the following phone, for example, [n] was realised as [ng] when followed by a velar sound such as [k]. Such type of feature spreading effects were identified with the help of spectrograms and EMA readings.

4.2 Duration changes

The total file duration of each audio file was compared with the duration of the corresponding EMA file and it was found that all the audio files were longer than their respective EMA files. The [sil] label at the end of each label file had a longer duration than the [sil] actually recorded for the EMA. Since the labelling was done originally on the acoustic data, the total duration of the file was equal to the duration of the audio file.

The sheet **Duration changes** in each excel file lists the name of each file and the difference between the duration of the audio and EMA files in seconds. The duration of each label file was adjusted to match the duration of the EMA file. No phone in the data set had duration smaller than 30ms.