

# EVALITA 2011 – Forced alignment task

## Test set

### Data description

The test set for the EVALITA 2011 Forced Alignment task consists of 89 wav files containing approximately 10 minutes of natural speech. Word level transcription for each file is provided inside the \*.wr\_ files.

Regarding the phonetic transcription, we let the participants choose if they want to employ their own automatic transcription system or if they prefer to have a dictionary. People using their own phonetic transcription systems will be awarded with bonus points during the final evaluation.

The dictionary that will be provided to people that will ask for it will contain the phonetic transcription of each word in the test set. The dictionary will contain only the EXPECTED phones sequence: there is no guarantee that all the phones in the dictionary transcription will be produced by the speaker. The reference phonetic transcription we will use for the final evaluation will not contain phones that were not actually pronounced. Should the speakers introduce sounds other than the ones expected from the transcription that are not simple garbage, these will not be contained in the transcription.

### Results submission

System results should consist of one automatically transcribed file for each turn following the format of the files in the training set.

For each word/phone, a single whitespace separated line should be generated containing

- The starting sample number of the word/phone
- The ending sample number of the word/phone
- The label of the word/phone

For example, the result of the automatic word alignment of the sequence “Salve a tutti” would be

```
0 5000 salve  
5000 15000 a  
15000 25000 tutti
```

while the result of the phonetic alignment would be

0 1000 s  
1000 2000 a  
2000 3000 l  
3000 4000 v  
4000 5000 e  
5000 15000 a  
15000 16000 t  
17000 18000 u  
18000 20000 t  
20000 25000 i

For the phonetic alignment task the results files extension should be \*.autophn while for the word alignment task the files extension should be \*.autowrd.

Final results should be submitted to [antonio.origlia@unina.it](mailto:antonio.origlia@unina.it) and to [cutugno@na.infn.it](mailto:cutugno@na.infn.it) no later than 14<sup>th</sup> October.