EVALITA 2009

Speech Recognition Task

Guidelines for Participants

Gianpaolo Coro
Abla srl, Viale Fulvio Testi, 7 Milano 20159
gianpaolo.coro@abla.it

Roberto Gretter, Marco Matassoni
FBK-irst, Povo via Sommarive 18, 38100 - Trento
{gretter,matasso}@fbk.eu

Task Description

In the Speech Recognition task, systems are required to recognize digit (from 0 to 9) sequences uttered in a speech signal. The task consists of two subtasks, and participants may choose to participate to any of them.

In the Clean Speech Digit Sequence Recognition Task, systems are required to recognize digit sequences in clean speech environment.

In the Noisy Speech Digit Sequence Recognition Task, systems are required to recognize digit sequences in noisy speech environment.

Corpus and Data Format

The Corpus consists of digit sequences coming from various Italian Corpora. It has been divided into training, development and test. Furthermore, it has been split into clean and noisy speech, according to the presence of many noise events in the signal or low SNR.
Audio files are sampled at 16kHz, 16 bit PCM, mono, Windows .wav format. For training and development set, transcription at word level will be provided in two separate text files, containing on each row the audio filename followed by the transcription.

```
E.g. training.txt:
clean00001.wav    1 3 5 6 4 0
clean00002.wav    3 5 7 9
…
noisy00022.wav    2 4 6 8 3


development.txt:
clean00101.wav    3 5 6 3 4
clean00102.wav    9 8 6 3
…
noisy00125.wav    7 9 6 8 3
```

Participants which need to enroll the ASR at finer level than phrase have to provide by themselves for the annotation. Participants are allowed to use training and development as they prefer.

Test data will be provided on September 10th, in the form of audio files and two lists of filenames (clean and noisy):

E.g. test_clean.txt:
```
00201.wav
00202.wav
…
```

Test results should be provided with the following format:

E.g. evalita09_SR_participant_type_run:
```
00201.wav  1 2 3 4 5
00202.wav  6 7 8 9 0
…
```

Submission of system results:

- Deadline: September 20th, 2009, midnight (GMT + 1 hour)
- System results have to be sent by e-mail to Roberto Gretter (gretter@fbk.eu), Marco Matassoni (matasso@fbk.eu) and Gianpaolo Coro (gianpaolo.coro@abla.it).
- Each participant has the possibility to submit a maximum of 2 runs for each task.
- System results will consist of a single data file and will have to be named as follows: `evalita09_SR_participant_type_run` where:
  - `participant` is the participant id;
  - `type` may be either *clean* or *noisy*;
  - `run` is an integer.

Evaluation Metrics

With respect to the results submitted by the participants, we will provide measurements of Word Accuracy and Sentence Accuracy.

Word Accuracy is defined as

$$WA = 100 - \frac{I + S + D}{N} \times 100$$

Referring to the automatic transcription,

- I is the number of inserted words
- S is the number of substitutions
- D is the number of the deletions
- N is the number of words in the reference

Sentence Accuracy: is defined as

$$SA = \frac{H}{M} \times 100$$

Referring to the automatic transcription,

- *H* is the number of sentences correctly recognized
- *M* is the number of sentences in the reference

The evaluation will be based on Minimum Edit Distance calculation between the transcription coming out from the recognizer and the orthographic annotation.

Contact Persons

Gianpaolo Coro (gianpaolo.coro@abla.it)
Roberto Gretter (gretter@fbk.eu)
Marco Matassoni (matasso@fbk.eu)