

EVALITA 2011

Forced Alignment on spontaneous speech task

Guidelines for participants

Francesco Cutugno¹, Antonio Origlia¹, Dino Seppi²

¹LUSI Lab, Dept. Of Physics, University of Naples “Federico II”, Italy

²ESAT, Katholieke Universiteit Leuven, Belgium

cutugno@na.infn.it, antonio.origlia@unina.it, dino.seppi@esat.kuleuven.be

Task Description

In this task, systems are required to align audio sequences of spoken dialogues (map task and difference test) to the provided relative transcriptions. The task has to be considered speaker independent. Two subtasks are defined, and applicants may choose to participate in any of them:

- Phone segmentation
- Word segmentation

Two modalities are allowed:

- Closed: only distributed data are allowed for training and tuning the system
- Open: the participant can use any type of data for system training, declaring and describing the proposed setup in the final report.

Materials

Training data: about 15 map task dialogues recorded by couples of speakers exhibiting a wide variety of Italian variants from the CLIPS corpus. Dialogues length ranges from 7/8 minutes to 15/20 minutes. It is up to participants to split these data in train and Dev subsets. For each dialogue the following files are provided:

- Full dialogue manually performed transcriptions;
- Single turn audio files: PCM-encoded mono WAV files (16KHz). Each file is referenced to turns into the full transcription by means of its name;
- Single turn phonetic labeling;
- Single turn word labeling

Test data: Unpublished dialogues from the material collected for the CLIPS corpus.

Data format

Information regarding speech data contained in each file can be extrapolated from the file name. File names are composed as follows:

- Corpus type: **DG** (Only dialogic data are contained in this set)
- Task type: **MT** (Map Task)
TD (Difference Test)
- Test ID: **A/B 0#** (A03, B05...)
- Dialect: Please refer to Table 1 for details regarding dialect coding
- Speaker ID: **_p1, _p2** for each test ID
- Role: **G/F** (Giver/Follower. Map task dialogues only)
- Turn number: **####** (#23, #254...)

Dialect	Label
Bari	B
Cagliari	C
Bergamo	D
Parma	E
Firenze	F
Genova	G
Catanzaro	H
Lecce	L
Milano	M
Napoli	N
Perugia	O
Palermo	P
Roma	R
Torino	T
Venezia	V

Table 1: Dialect codes

More details regarding each speaker (Gender, age, birthplace...) can be found in the included dialogue description files (TXT) along with the full dialogue transcription.

Transcription files are encoded as ASCII files containing an N*3 table where N is the number of segments (words or phones) in the file. Each row of the table reports the starting sample of the corresponding segment, its end sample and its label.

In Table 2 the set of SAMPA symbols used in Phn files is reported. Due to the difficulty of finding a marker between two vowels, the annotation rule was not to split the occurrence of this situation.

a	k	m	ja
e	g	n	je
E	ts	J	jo
i	dz	r	ju
o	tS	l	oj
O	dZ	L	wa
u	f	__	we
p	v	aj	wi
b	s	aw	wo
t	z	ej	# (garbage)
d	S	ew	

Table 2: Phn files symbols set

Symbols other than words used in the Wrd files are reported in Table 3

#	Garbage
<sp>	Short pause
<lp>	Long pause
<P>	Medium/long pause with discourse interruption
<ehm>, <eeh>...	Filled pauses
word<vv> <vv>word	Filled pauses with vowel lengthening (allora<aa>, <ee>eccolo)
<cc>word word<cc>	Filled pauses with consonant lengthening (<ss>senti, non<nn>...)
wo_rd	Internal interruptions (mon_tato)

Table 3: Wrd files special symbols set

Submission of system results

Deadline: October the 14th, 2011, midnight (GMT+1 hour)

Results should be submitted by email to Antonio Origlia (antonio.origlia@unina.it)

System results should consist of one automatically transcribed file for each turn following the format of the original files. For the phone alignment task the files extension should be *.autophn while for the word alignment task the files extension should be *.autowrd

Evaluation metrics

The evaluation is based on Unit Boundary Positioning Accuracy. The evaluation methodology will follow the standard described in the documentation of the NIST SCLite evaluation tool. The SCLite tool itself will be used as scorer.

Contact Person

Antonio Origlia: antonio.origlia@unina.it

References

R. Savy and F. Cutugno, "Clips: diatopic, diamesic and diaphasic variations of spoken italian," in Proc. of Corpus Linguistics Conference, 2009.
[Online]. Available: <http://ucrel.lancs.ac.uk/publications/cl2009/>, ID:213

Unit Boundary Positioning Accuracy (SCLite documentation).
[Online]. Available: <http://www.icsi.berkeley.edu/Speech/docs/sctk-1.2/sclite.htm>

SCLite tool (included in the NIST SCTk suite)
[Online]. Available: www.itl.nist.gov/iad/mig/tools