The SPPAS participation to Evalita 2011

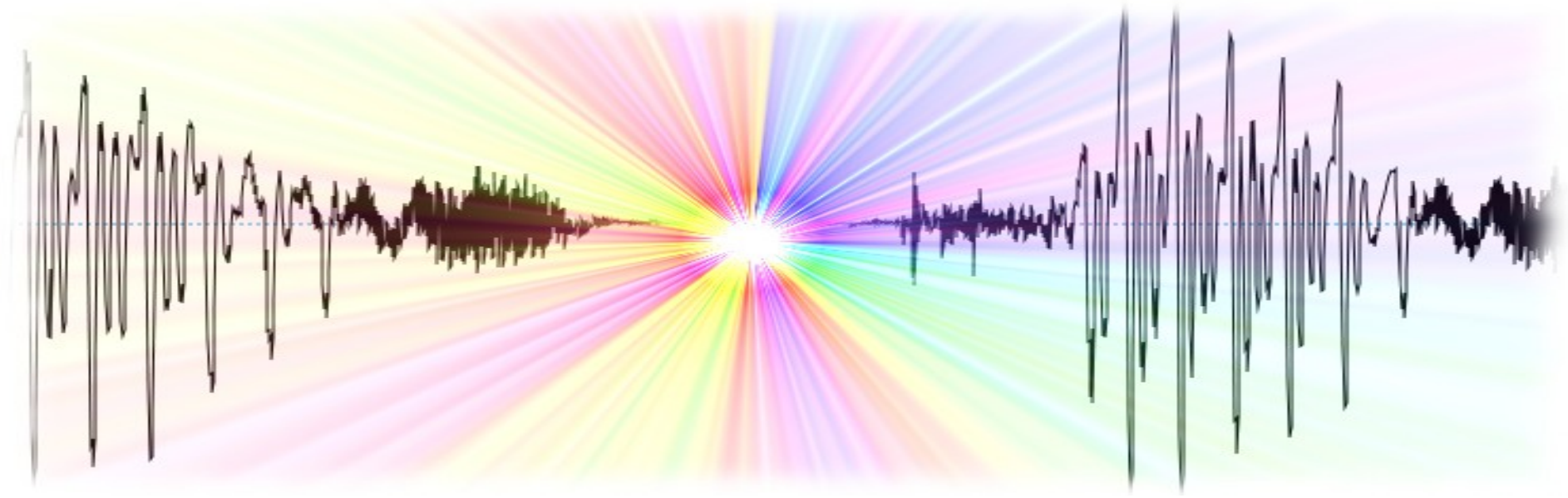
Brigitte Bigi

UMR 6057 CNRS
PAROLE ET LANGAGE

cnrs
dépasser les frontières

Aix*Marseille université

ma che cos' è SPPAS?

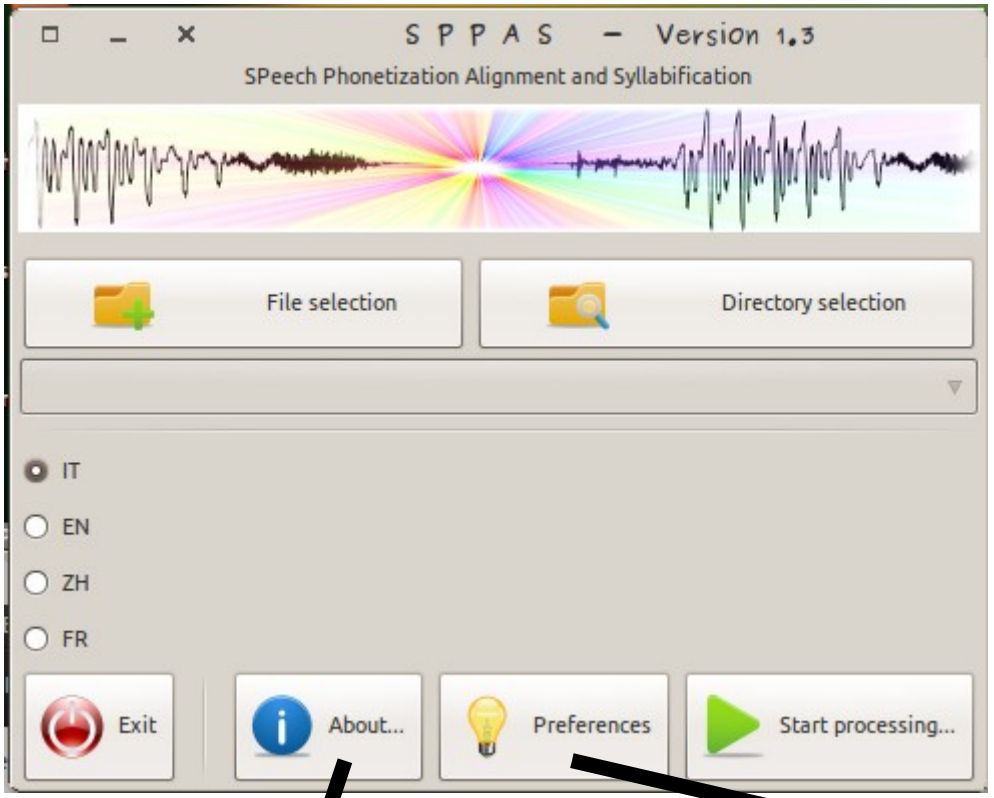


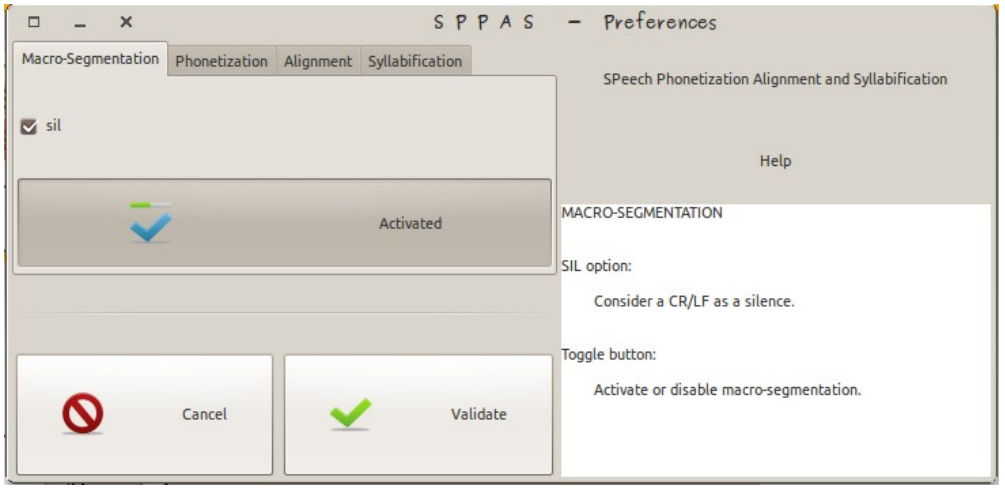
SPeech Phonetization Alignment and Syllabification

Main description

- A new tool to produce automatically annotations which includes utterance, word, syllabic and phonemic segmentations from a recorded speech sound and its transcription.
- Language-independent.
- Currently designed for French, English, Italian and Chinese and there is an easy way to add other languages.
- Distributed under GPL license.

SPPAS Screenshots





Evalita task

- ”Forced Alignment on Spontaneous Speech”:
  - ✓ Phone segmentation;
  - ✓ Word segmentation.
- Data: ✓ Closed; · Opened.
- Dialogues, map-tasks:
  - 3h30 speech;
  - 15% phones: pauses, filled-pauses, garbage.
- SPPAS Forced-Alignment is 2 sub-tasks:
  - phonetization + alignment

Phonetization

- The process of representing sounds by phonetic signs.
- There are two general ways to construct a phonetization process:
  - rule based systems (with rules based on inference approaches or proposed by expert linguists);
  - dictionary based solutions which consist of storing a maximum of phonological knowledge in a lexicon.

SPPAS phonetization

- SPPAS uses the dictionary-based approach.
- The phonetization is the equivalent of a sequence of dictionary-look-ups:
  - Input transcription needs to be word-segmented
  - It is supposed that all words of the transcription are mentioned in the pronunciation dictionary.
- A specific phone to represent filled pauses.

Dictionary

- The dictionary contains a set of possible pronouciations of words, including accents as *"perchè"* pronounced as `/b e r k e/`, and reduction phenomena as `/p e k/`.
- Made of:
  - 390k words;
  - 5k variants.

```
264377 PESAVO [PESAVO] p e s a v o
264378 PESCA [PESCA] p E s k a
264379 PESCA(2) [PESCA] p e s k a
264380 PESCADOR [PESCADOR] p e s k a d o r
264381 PESCAGGI [PESCAGGI] p e s k a dZ i
264382 PESCAGGIO [PESCAGGIO] p e s k a dZ o
264383 PESCAI [PESCAI] p e s k a i
264384 PESCAIA [PESCAIA] p e s k a ja
```

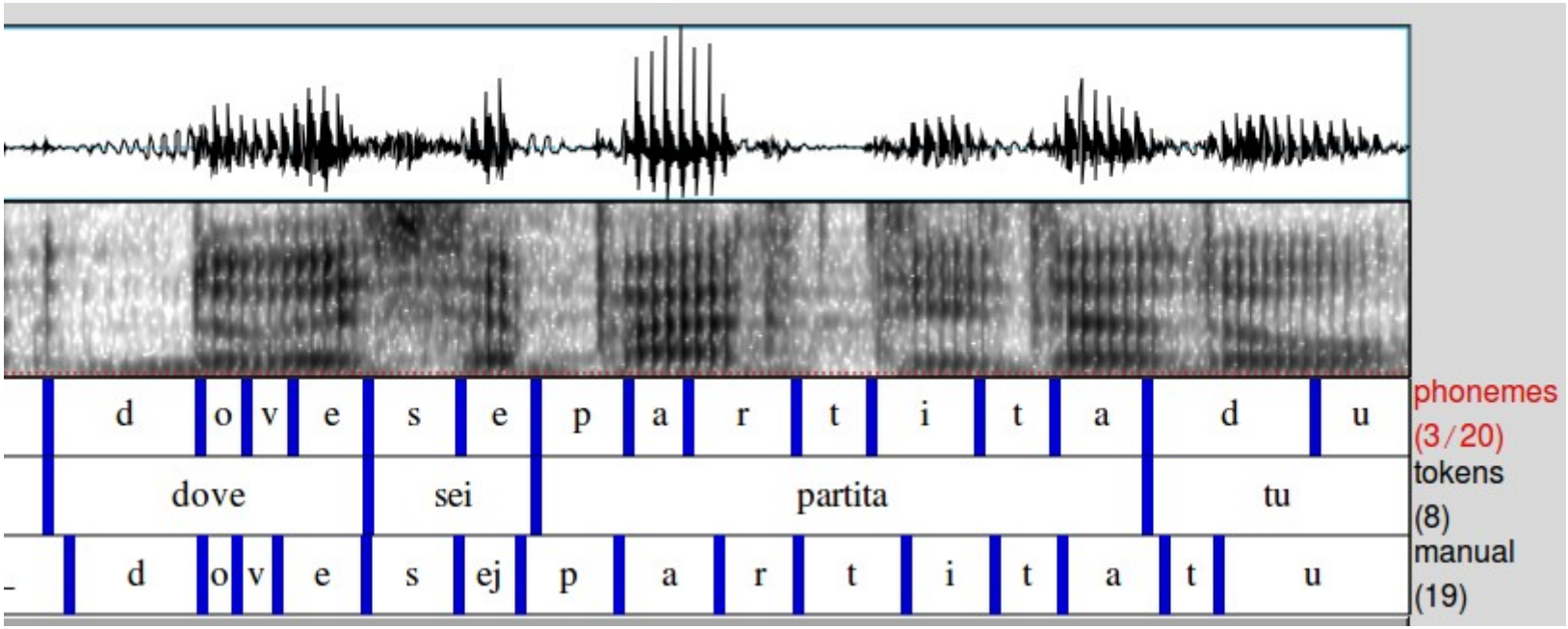
- From:
  - Festival + Evalita training corpus.
- Manual corrections

Phonetization: variants

- No rules are applied:
  - all possibilities are proposed to the aligner.
- For example, the sentence "*del mio frigorifero*" will produce the following list of words with associated pronunciations:
  - `d.e.l | d.E.l`
  - `m.jo | m.i`
  - `f.r.i.g.o.r.i.f.e.r.o | f.r.i.g.o.r.i.f.E.r.o | f.r.i.g.o.r.i.f.e.r | f.r.i.g.o.r.i.f.e.r.O`

Alignment

- A time-matching between a given speech utterance along with a phonetic representation of the utterance.



SPPAS Alignment

- Alignment in SPPAS is based on the Julius Speech Recognition Engine (SRE):
  - A finite state grammar that describes sentence patterns to be recognized;
  - An acoustic model.

- The alignment task is a 2-steps process:
  - the first one choose the phonetization;
  - the second one perform the segmentation.

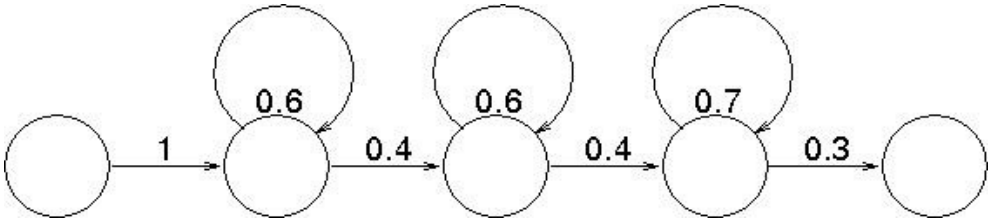
Grammar

- A grammar: constraints on what the SRE can expect as input. It is a list of words that the SRE listens for. Each word has a set of associated list of phonemes.

Dictionary	Grammar
0 w_0 d e l 0 w_0 d E l 1 w_1 m jo 1 w_1 m i 2 w_2 f r i g o r i f e r o 2 w_2 f r i g o r i f E r o 2 w_2 f r i g o r i f e r 2 w_2 f r i g o r i f e r O	0 2 1 0 1 1 1 2 0 0 2 0 3 0 0 3 -1 -1 1 0

Acoustic Model

- HMM, 5-states
- Triphones
- Trained with HTK:
  - from the proposed phonetized transcription, without using the phonetic time-alignment;
  - using 16 bits, 16000 hz wav files;
  - Mel-frequency cepstrum coefficients (MFCC) along with their first and second derivatives were extracted from the speech in the standard way: MFCC_D_N_Z_0.



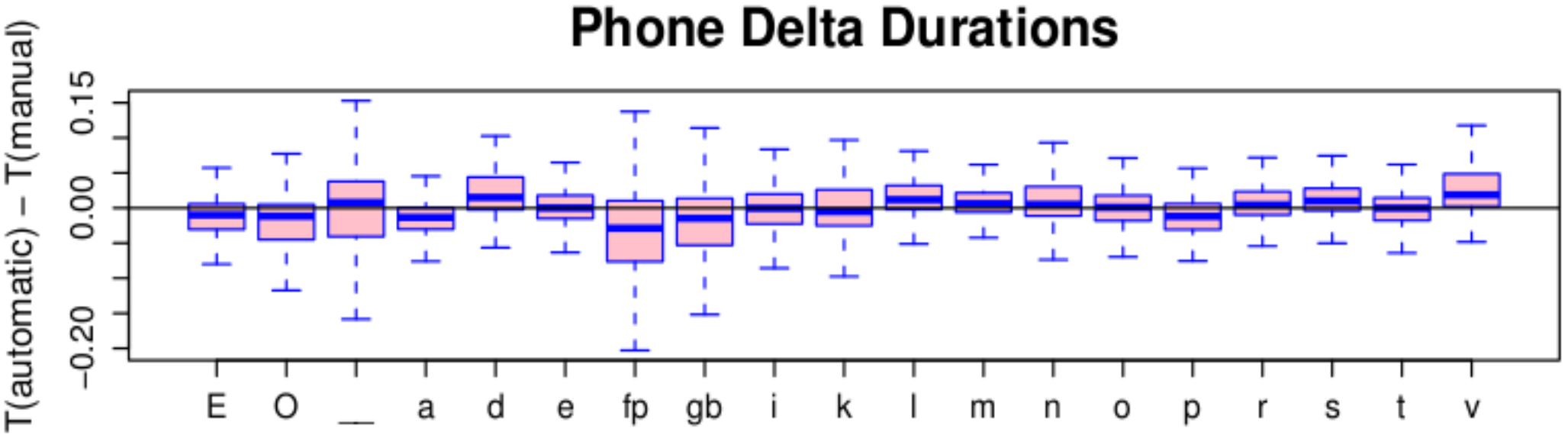
Results

- Development corpus:
  - 200 utterances; 12 min 04 sec.
  - 2373 words, 6282 phonemes, including:
    - 689 “_” (pauses);
    - 246 “#” (garbage);
    - both represent 14.88%.
- Evaluate separately:
  - Alignment;
  - Phonetization obtained after the alignment.

Results alignment

- The availability of our system to align the good phoneme sequence.
- On the basis of the *manual phonetization.*
- Sclite using the time-alignment option :
  - a correct rate of 89.8%, with 7.6% substitutions, 2.6% deletions and 2.6% insertions.

Results alignment



- Pauses, filled pauses and garbages: greatest ranges
- Vowels: automatic shorter than manual
- Consonants: automatic higher than manual

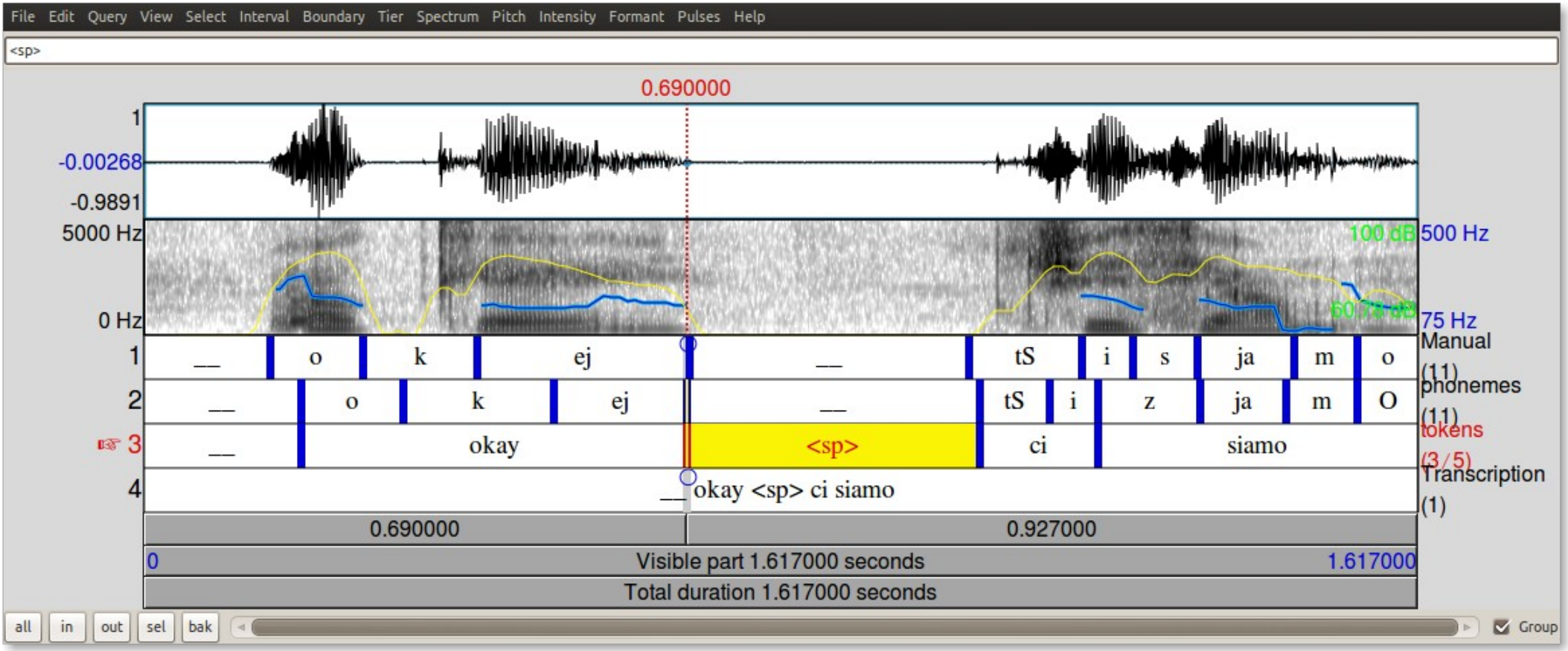
Results phonetization

- The availability of our system to propose the expected phoneme sequence.
- Sclite *without* using the time-alignment option :
  - a correct rate of 89.5%, with 8.1% substitutions, 2.3% deletions and 6.9% insertions.
- Most frequent errors are due to the garbage manual annotation. Example: *bravissimo a questo*
  - Automatic: `b r a v i s i m o a k we s t o`
  - Manual: `b r # s # k we s t o`
    - 5 insertions, 2 substitutions!

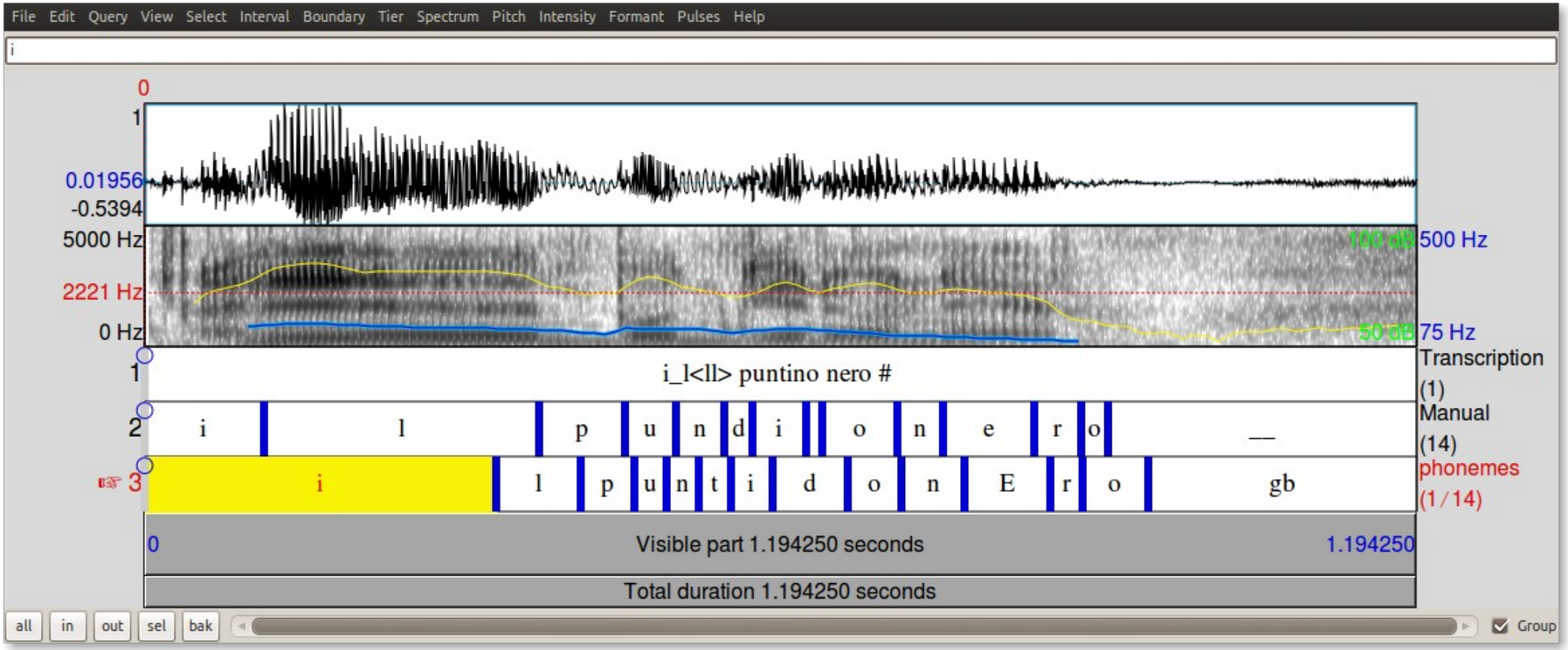
SPPAS final results

- Official results estimated using sclite:
  - 88.4% good phoneme alignments:
    - This score contains both phonetizations and alignments errors.
  - 96.7% good word alignments.

Example


File Edit Query View Select Interval Boundary Tier Spectrum Pitch Intensity Formant Pulses Help
<sp>
0.690000
1
-0.00268
-0.9891
5000 Hz
0 Hz
100 dB
500 Hz
75 Hz
1 __ o k ej __ tS i s ja m o Manual (11)
2 __ o k ej __ tS i z ja m O phonemes (11)
3 __ okay <sp> ci siamo tokens (3/5)
4 __ okay <sp> ci siamo Transcription (1)
0.690000
0.927000
0
Visible part 1.617000 seconds
1.617000
Total duration 1.617000 seconds
all in out sel bak
Group

Example


File Edit Query View Select Interval Boundary Tier Spectrum Pitch Intensity Formant Pulses Help
i
0
1
0.01956
-0.5394
5000 Hz
2221 Hz
0 Hz
100 dB
500 Hz
50 dB
75 Hz
1
i_l<ll> puntino nero #
Transcription
(1)
2
i l p u n d i o n e r o __
Manual
(14)
3
i l p u n t i d o n E r o gb
phonemes
(1/14)
0
Visible part 1.194250 seconds
1.194250
Total duration 1.194250 seconds
all in out sel bak
Group

Conclusion

- SPPAS: a tool to perform the forced-alignment task during the Evalita 2011 campaign, on Italian map-task dialogues.
- SPPAS was not specifically devoted to Italian: it can deal with various languages: French, English, Chinese.
- http://www.lpl-aix.fr/~bigi/sppas/