What's new in SPPAS 1.5?

Brigitte Bigi, Daniel Hirst

brigitte.bigi@lpl-aix.fr, daniel.hirst@lpl-aix.fr

Abstract

During Speech Prosody 2012, we presented SPPAS, a tool to automatically produce annotations which include utterance, word, syllabic and phonemic segmentations from a recorded speech sound and its transcription. SPPAS is specifically designed to be used directly by linguists.

Automatic Annotations: Segmentation

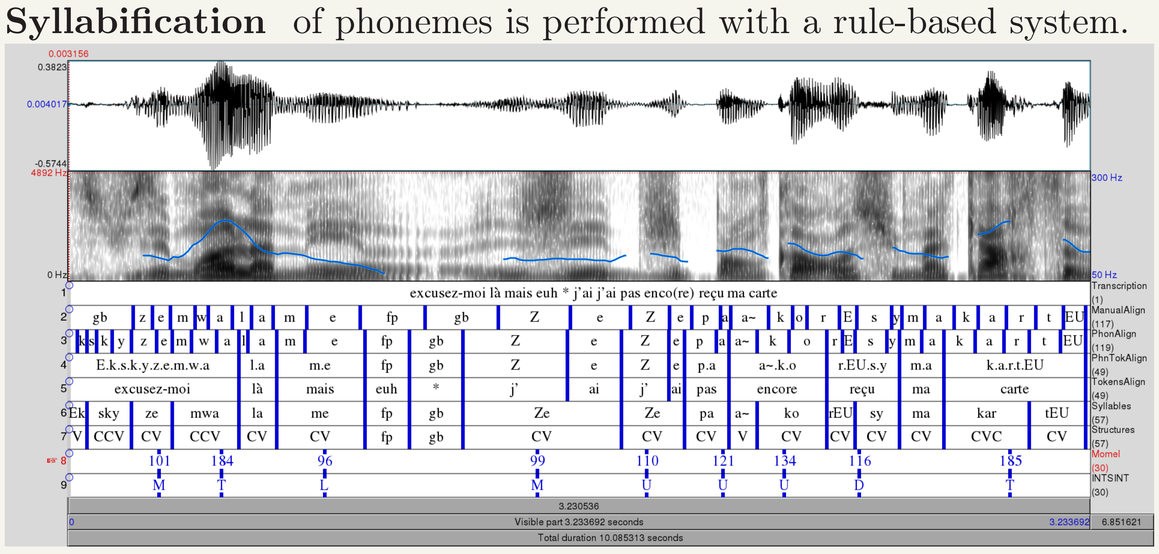
Inter-Pausal Units segmentation consists in aligning the macro-units of a document (based on their transcription) with the corresponding sound. IPUs Segmentation annotation performs a simple silence detection if no transcription is available.

Tokenization is the process of segmenting a text into tokens. SPPAS implements a generic approach for text normalisation, in view of developping a multi-purpose multi-lingual text corpus.

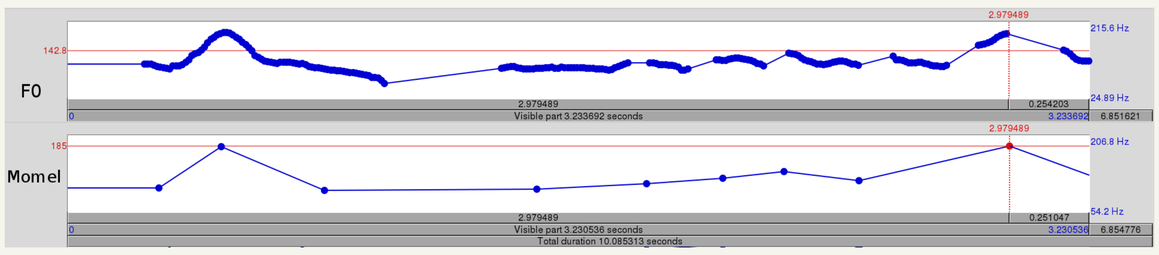
Phonetisation is the process of representing sounds with phonetic signs. The phonetisation is the equivalent of a sequence of dictionary look-ups. SPPAS implements a language-independent algorithm to phonetise unknown words.

Phonetic Alignment consists in a time-matching between a given speech utterance and a phonetic representation of the utterance. For each utterance, the orthographic and phonetic transcriptions are used. SPPAS call the Julius CSR engine to perform alignment.

Syllabification of phonemes is performed with a rule-based system.



Automatic Annotations: Prosody



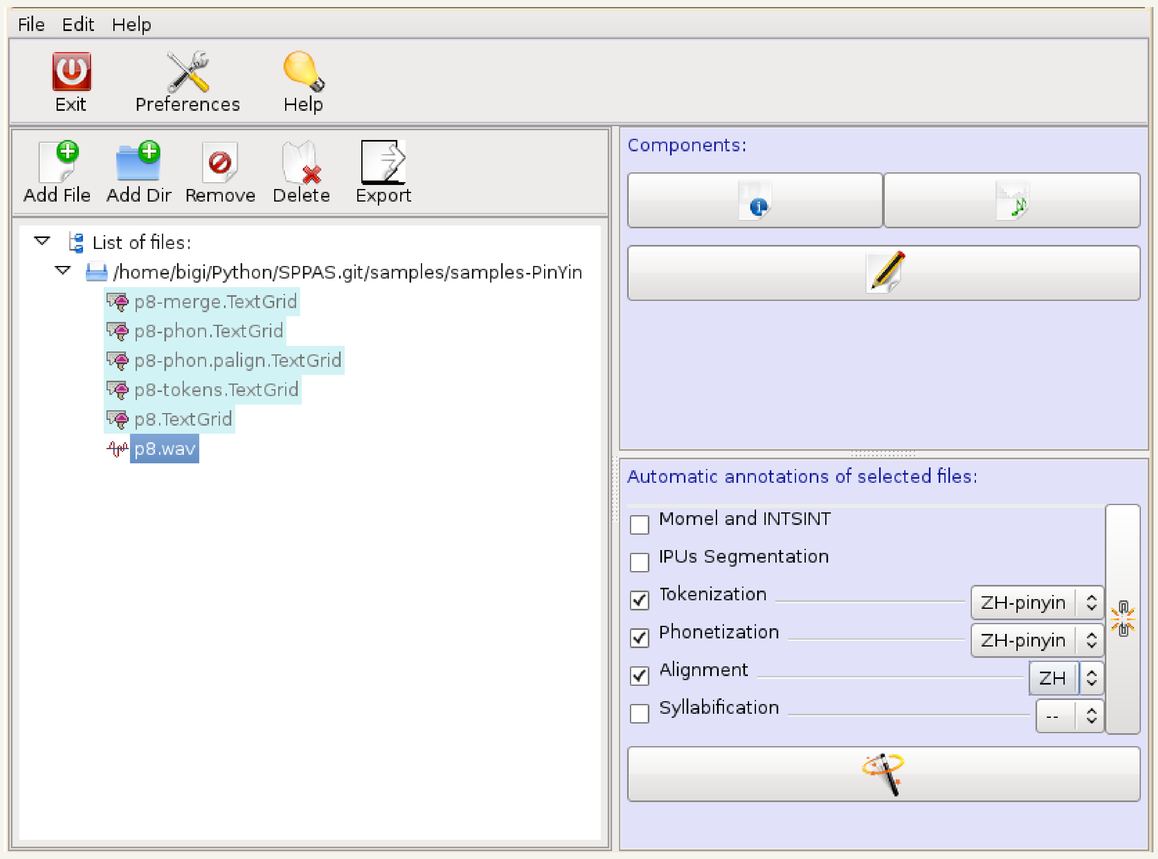
Momel (modelling melody): automatic modelling of fundamental frequency (F0) curves, using a technique called assymmetric modalquadratic regression. This technique makes it possible by an appropriate choice of parameters to factor an F0 curve into two components:

1. a macroprosodic component represented by a a quadratic spline function defined by a sequence of target points <ms, hz>.
2. a microprosodic component represented by the ratio of each point on the F0 curve to the corresponding point on the quadratic spline function.

Since several different techniques of F0 extraction are possible, Momel requires a file containing the F0 values detected from the signal.

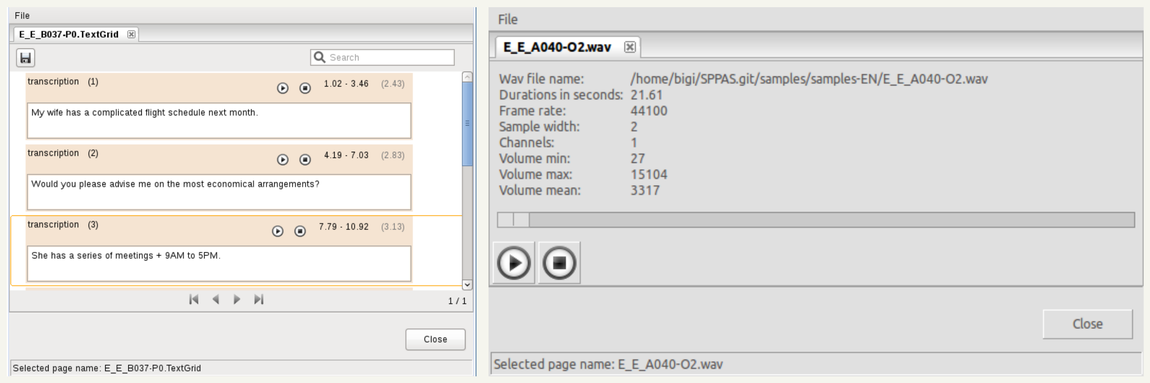
INTSINT: Encoding of F0 target points using T (Top), M (Mid), B (Bottom), H (Higher), S (Same), L (Lower), U (Upstepped), D (Downstepped) each one of which characterises a point on the fundamental frequency curve. The rationale behind the INTSINT system is that the F0 values of pitch targets are programmed in one of two ways: either as absolute tones T, M, B which are assumed to refer to the speaker's overall pitch range (within the current Intonation Unit), or as relative tones H, S, L, U, D assumed to refer only to the value of the preceding target point. A distinction is made between non-iterative H, S, L and iterative U, D relative tones since in a number of descriptions it appears that iterative raising or lowering uses a smaller F0 interval than non-iterative raising or lowering.

SPPAS 1.5

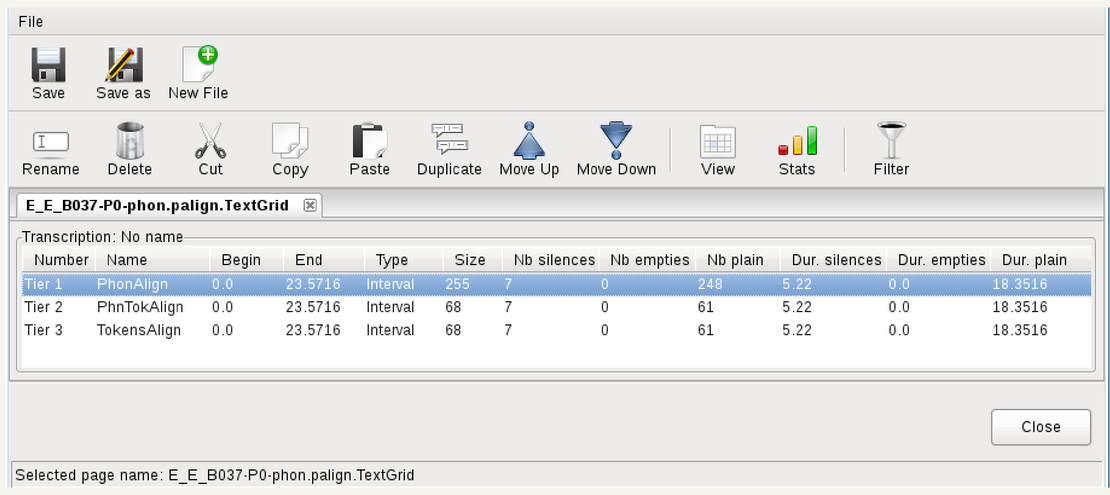


http://www.lpl-aix.fr/~bigi/sppas/

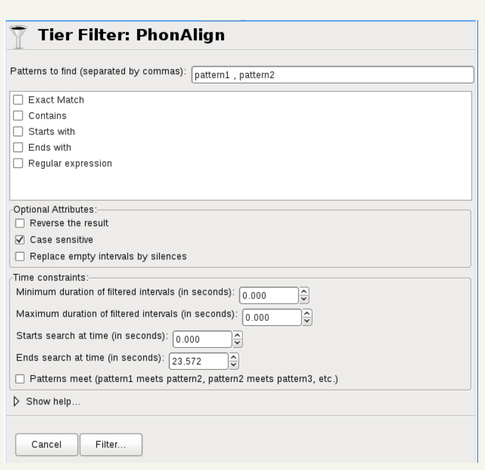
Components: Transcribe and Player

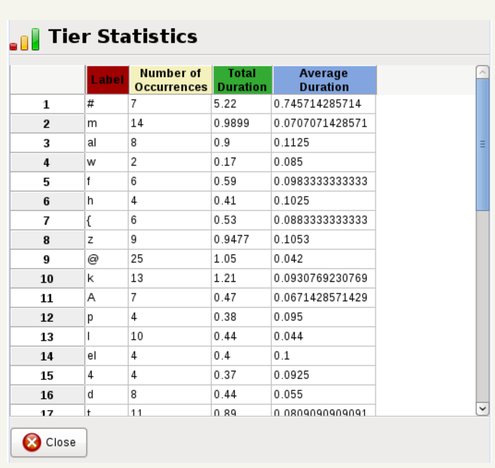


Components: Information and Requests



To get information, modify and request annotated files. It allows the user to manage annotated files and the tiers of these files: rename, delete, cut, copy, paste duplicate, move up, move down, view, prints elementary statistics, filter annotated data.





Acknowledgements

This research was partially supported by the Ortolang project. The support is gratefully acknowledged. http://www.ortolang.fr.