

What's new in SPPAS 1.5?

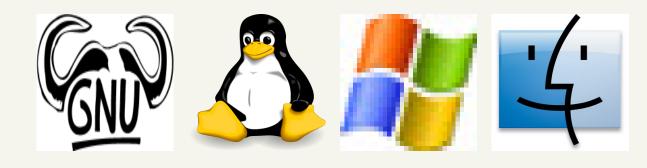
Brigitte Bigi, Daniel Hirst



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

Abstract

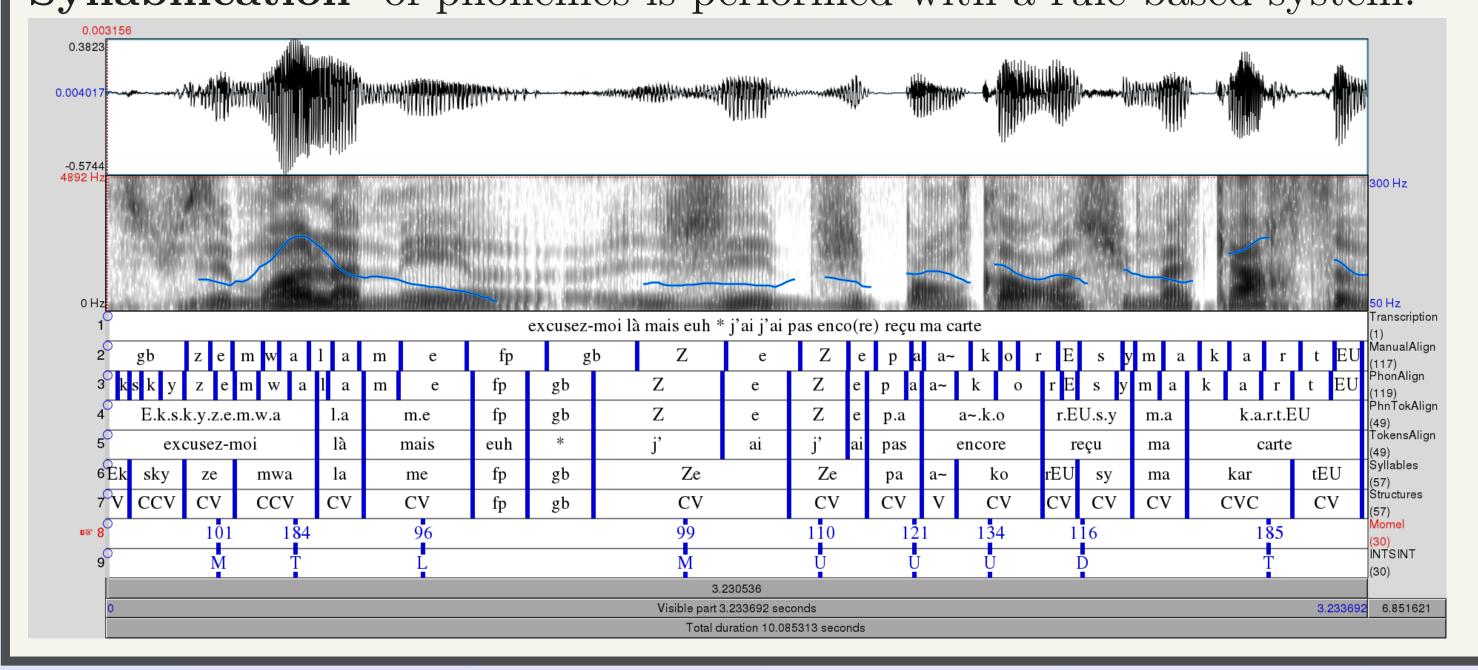
During Speech Prosody 2012, we presented SPPAS, a tool to automatically produce annotations which include utterance, word, syllabic and phonemic segmentations from a recorded speech sound and its transcription. SPPAS is specifically designed to be used directly by linguists.

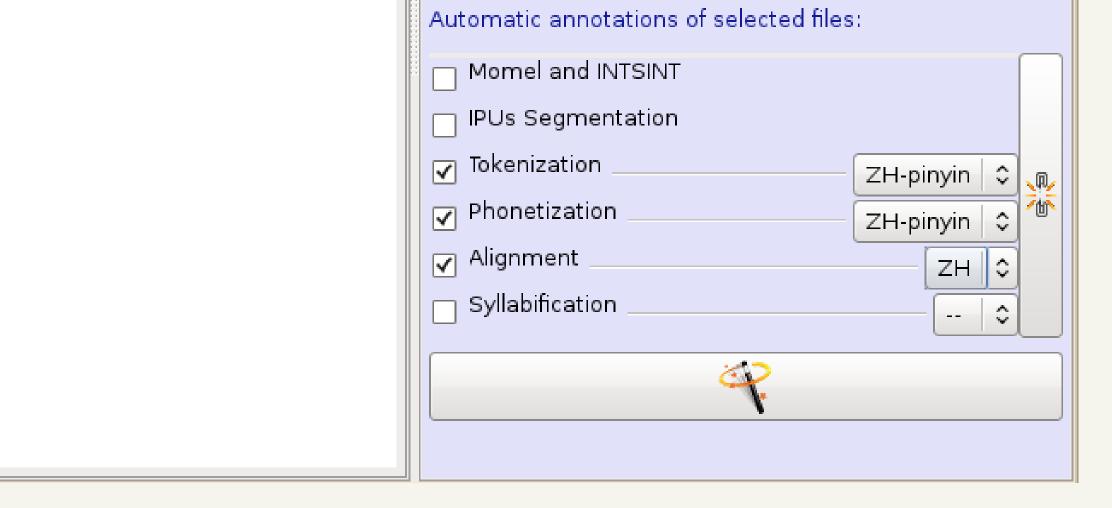


Automatic Annotations: Segmentation

Inter-Pausal Units segmentation consists in aligning the macrounits 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.

SPPAS 1.5 File Edit Help U Preferences Exit Components: 0 Add File Add Dir Remove Delete Export O 🗢 🚼 List of files: Image: Antipathon / SPPAS.git/samples/samples-PinYin 🏹 p8-merge.TextGrid 쪆 p8-phon.Text Grid 👼 p8-phon.palign.TextGrid 쪑 p8-tokens.TextGrid 🍖 p8.TextGrid

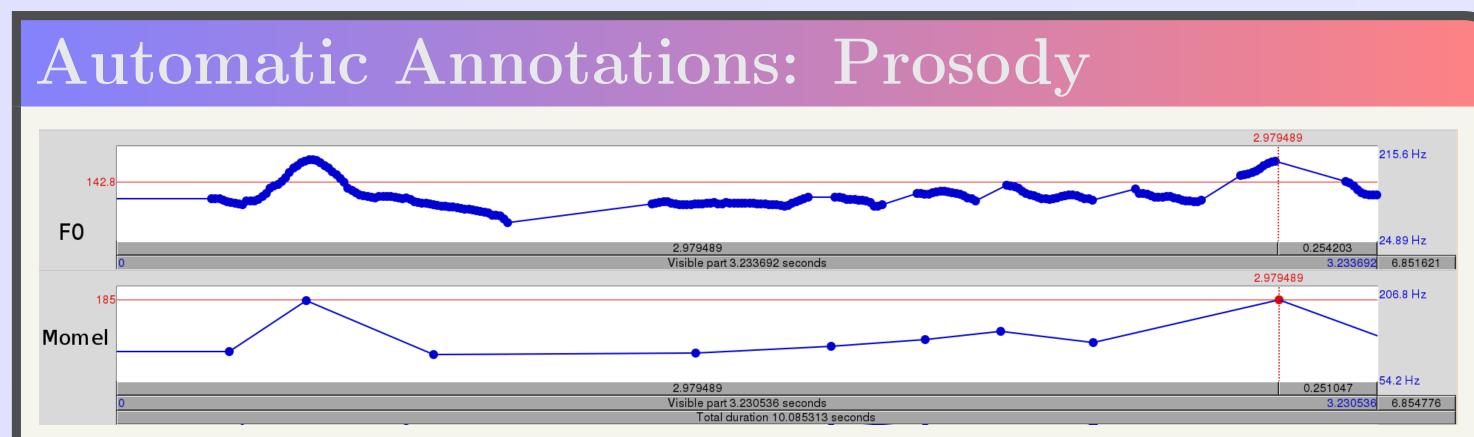




http://www.lpl-aix.fr/~bigi/sppas/

Components: Transcribe and Player

File		File		
E_E_B037-P0.TextGrid 🗷				
	Q Search	E_E_A040-O2.wav 🙁		
transcription (1)	 1.02 · 3.46 (2.43) 	Wav file name: /home/bigi/ Durations in seconds: 21.61	/SPPAS.git/samples/samples-EN/E_E_A040-O2.wav	
My wife has a complicated flight schedule next month.	≡	Frame rate: 44100 Sample width: 2		
transcription (2)	● 4.19 · 7.03 (2.83)	Channels: 1 Volume min: 27 Volume max: 15104		
Would you please advise me on the most economical arra	angements?	Volume mean: 3317		
transcription (3)	7.79 · 10.92 (3.13)			
She has a series of meetings + 9AM to 5PM.				
ia a de de	1/1			
	Close			Close
Selected page name: E_E_B037-P0.TextGrid		Selected page name: E_E_A040-O2.	.wav	



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:

a macroprosodic component represented by a a quadratic spline function defined by a sequence of target points $\langle ms, hz \rangle$.

2. a microprosodic component represented by the ratio of each point on the F0 curve to the corresponding point on the quadratic spline function.

Components: Information and Requests

I Rename	Delete	Cut	Сору	Paste	Duplicate	Move Up	Move Down	Uiew	 Stats	Filter	
	7-P0-phon.pa										
	ion: No name-	0									
	Name	Begin	End	Туре	Size	Nb siler	ces Nb emptie	es Nb plain	Dur. siler	ices Dur. emp	oties Dur. plain
Tier 1	PhonAlign	0.0	23.5716	Interval	255	7	0	248	5.22	0.0	18.3516
Tier 2	PhnTokAlign	0.0	23.5716	Interval	68	7	0	61	5.22	0.0	18.3516
Tier 3	TokensAlign	0.0	23.5716	Interval	68	7	0	61	5.22	0.0	18.3516

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.

Tier Filter: PhonAlign	Tier Statistics
Patterns to find (separated by commas): pattern1 , pattern2	Label Number of Total Average Occurrences Duration

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.

Exact Match	1	#	7	5.22	0.745714285714
Contains		-	14	0.9899	0.0707071428571
Starts with	2		14		
Ends with	3	al al	8	0.9	0.1125
Regular expression	4	k w	2	0.17	0.085
	5	i f	6	0.59	0.09833333333333
ional Attributes:	6	i h	4	0.41	0.1025
Reverse the result	7	′ {	6	0.53	0.08833333333333
Case sensitive	8	z	9	0.9477	0.1053
Replace empty intervals by silences	9	@	25	1.05	0.042
e constraints:	1	0 k	13	1.21	0.0930769230769
nimum duration of filtered intervals (in seconds): 0.000	1	1 A	7	0.47	0.0671428571429
ximum duration of filtered intervals (in seconds): 0.000	1	2 p	4	0.38	0.095
arts search at time (in seconds): 0.000	1	3	10	0.44	0.044
ds search at time (in seconds):	1	4 el	4	0.4	0.1
Patterns meet (pattern1 meets pattern2, pattern2 meets pattern3, etc.)	1	5 4	4	0.37	0.0925
Show help	1	6 d	8	0.44	0.055
Show help	1	7 t	11	0.89	0 0809090909091

Acknowledgements

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