1An integrated tool for (macro)syntax-intonation correlation analysis

Philippe Martin

UMR 7110, LLF, UFRL, Université Paris Diderot, ODG, rue Albert Einstein, 75013 Paris, France
philippe.martin@linguist.univ-paris-diderot.fr

Abstract

Many efforts are presently made to elaborate large corpora of spontaneous speech (PFC, CFPP2000, C-PROM, ESLO, PFC, Rhapsodie, ORFEO to quote a few), in order to offer the research community large databases that could be used in many aspects of linguistic research. I introduce here new functions of the WinPitch software addressing two aspects of oral corpus analysis:

1) Data mining functions involving specific speech unit (such as conjunctions, weak verbs, etc.) to retrieve rapidly and efficiently their occurrences and their context, displaying automatically the corresponding speech segments together with their acoustical analysis (F0 curve, spectrogram, etc.)

2) Tools enabling the correction of pitch curves resulting from adverse recording conditions, in order to obtain reliable F0 data for further processing (statistical analysis, automatic annotation of sentence intonation, etc.).

Index Terms: spontaneous speech, fundamental frequency, intonation transcription, concordancer.

1. Introduction

A lot of interest is presently devoted to the linguistic analysis of non-prepared speech, and in particular to the prosodic correlates of syntactic and macrosyntactic units. In this type of research, it is assumed that prosodic events help the listener to dynamically reconstruct the prosodic structure intended by the speaker, and eventually allow to infer the syntactic organization of the sentence with which the prosodic structure may be congruent or not.

To investigate this process, it seems at first that patient and meticulous examination of data would be required. Say for example that we want to know about the prosodic correlates of the occurrences of the conjunction "*parce que*" (because) in a set of spontaneous recordings of French.

Instead of listening to hours of recordings to retrieve pronounced occurrences of the key word, we would attempt to retrieve "*parce que*" in all the available text transcriptions of the recordings, and find in a second stage the corresponding speech segments in order to analyze their prosodic properties. This task would be further facilitated if the transcription is aligned, i.e. if bidirectional pointers between text segments and corresponding speech segments have been implemented. This would enable the easy retrieval of every occurrence of the appropriate speech segment from a text selection.

Still, most of the tools available today stop at this stage, even if concordancer of transcribed text items are readily available, listing all occurrences of the search item with together with its left and right contexts.

The new function implemented in WinPitch goes a little bit further by providing the following functions to allow the user to efficiently and rapidly examine a large number of data with a minimum of manipulations:

1. Generation of a text transcription from alignment files in various largely used formats (Praat textgrid, Transcriber trs, C-Oral Rom xml. Necte xml, CRF alg, etc.);

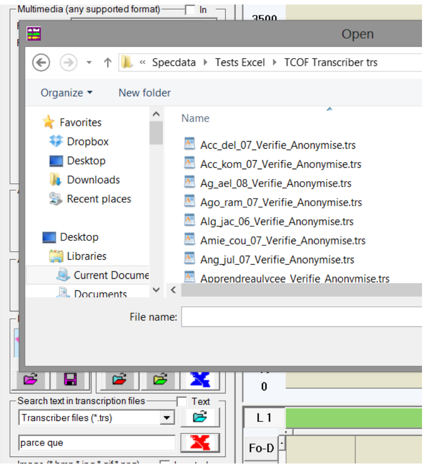
2. Concordancer: generation of a list of occurrences of the search word, with its left and right contexts. This list is automatically created in Excel format;

3. Automatic retrieval of the search word occurrence in a context selected by the user on the Excel table generated in step 2, with a single mouse click.

4. Extraction of the corresponding speech segment from the proper sound file, played back with all relevant acoustical data displayed (spectrogram, fundamental frequency F0, intensity and duration curves).

2. Integrated concordancer

Figures 1 to 4 illustrate the details of the operations involved. In Fig. 1 The user enters the key word "*parce que*", selects an appropriate alignment format (Transcriber trs in this example), and clicks on any of the file names stored in a common directory. This directory should contain all the alignment files of interest, together with their corresponding sound files. In the case of Praat textgrid files, the corresponding sound files must have the same name as their textgrid counterpart, as Praat textgrid files do not contain any reference to their corresponding speech file.



Specdata › Tests Excel › TCOF Transcriber trs

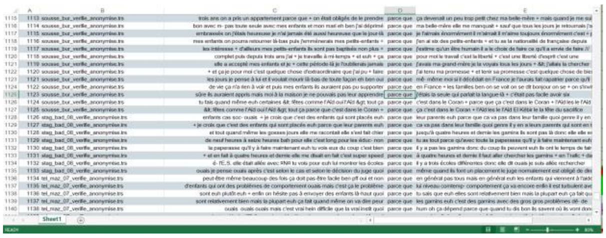
Acc_del_07_Verifie_Anonymise.trs
Acc_kom_07_Verifie_Anonymise.trs
Ag_ael_08_Verifie_Anonymise.trs
Ago_ram_07_Verifie_Anonymise.trs
Alg_jac_06_Verifie_Anonymise.trs
Amie_cou_07_Verifie_Anonymise.trs
Ang_jul_07_Verifie_Anonymise.trs
Apprendreaulycee_Verifie_Anonymise.trs

File name:

Search text in transcription files
Transcriber files (*.trs)
parce que

Figure 1. Entering the key word "parce que" and selecting a Transcriber files in a directory containing all files of interest.

An Excel table listing all found occurrences of the key word is immediately generated (Fig. 2). This operation is very fast, in the example of *parce* que, the completion takes less than one second to scan 104 files giving 1194 occurrences.

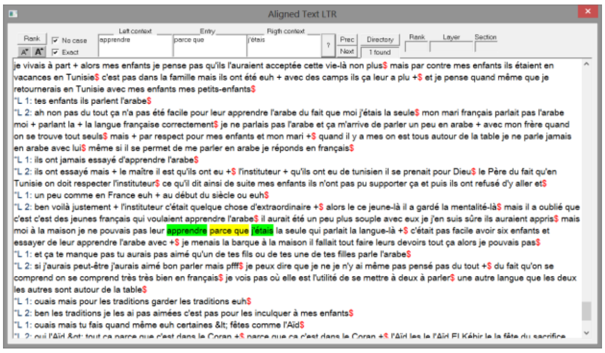


	A	B	C	D	E
1115	1113	sousse_bur_verifie_anonymise.trs	trois ans on a pris un appartement parce que + on était obligés de le prendre	parce que	ça devenait un peu trop petit chez ma belle-mère + mais quand je me su
1116	1114	sousse_bur_verifie_anonymise.trs	bon avec m- pas toute seule avec mes enfants et mon mari eh ben j'ai déprimé	parce que	ma belle-mère elle me manquait + sauf que tous les jours je retournais j'a
1117	1115	sousse_bur_verifie_anonymise.trs	embrassée on j'étais heureuse je n'ai jamais été aussi heureuse que le jour-là	parce que	je l'aimais énormément il m'aimait il m'aime toujours énormément c'est + j
1118	1116	sousse_bur_verifie_anonymise.trs	mes enfants on pourra retourner là-bas puis j'emmènerais mes petits-enfants +	parce que	j'en ai six des petits-enfants + et tu as la nationalité de française depuis
1119	1117	sousse_bur_verifie_anonymise.trs	les intéresse + d'ailleurs mes petits-enfants ils sont pas baptisés non plus +	parce que	j'estime qu'un être humain il a le choix de faire ce qu'il a envie de faire //
1120	1118	sousse_bur_verifie_anonymise.trs	complet puis depuis trois ans j'ai + je travaille à mi-temps + et euh + ça	parce que	pour moi le travail c'est la liberté + c'est une liberté d'esprit c'est une
1121	1119	sousse_bur_verifie_anonymise.trs	elle a accepté mes enfants et je + cette période-là je l'oublierais jamais	parce que	j'avais ma grand-mère je la voyais tous les jours + &lt; j'allais la chercher
1122	1120	sousse_bur_verifie_anonymise.trs	+ et ça je pour moi c'est quelque chose d'extraordinaire que j'ai pu + faire	parce que	j'ai tenu ma promesse + et tenir sa promesse c'est quelque chose de bie
1123	1121	sousse_bur_verifie_anonymise.trs	les jours je pense à lui et il voulait mourir là-bas de toute façon eh ben oui	parce que	mê- même moi si il décédait en France je l'aurais fait rapatrier parce qu'il
1124	1122	sousse_bur_verifie_anonymise.trs	de vie ça n'a rien à voir et puis mes enfants ils auraient pas pu supporter	parce que	en France + les familles ben on se voit on se dit bonjour on se + on s'invi
1125	1123	sousse_bur_verifie_anonymise.trs	sûre ils auraient appris mais moi à la maison je ne pouvais pas leur apprendre	parce que	j'étais la seule qui parlait la langue-là + c'était pas facile avoir six
1126	1124	sousse_bur_verifie_anonymise.trs	tu fais quand même euh certaines &lt; fêtes comme l'Aïd oui l'Aïd &gt; tout ça	parce que	c'est dans le Coran + parce que ça c'est dans le Coran + l'Aïd les le l'Aïd
1127	1125	sousse_bur_verifie_anonymise.trs	&lt; fêtes comme l'Aïd oui l'Aïd &gt; tout ça parce que c'est dans le Coran +	parce que	ça c'est dans le Coran + l'Aïd les le l'Aïd El Kébir le la fête du sacrifice
1128	1126	stag_bad_08_verifie_anonymise.trs	enfants cas soc- ouais + je crois que c'est des enfants qui sont placés euh	parce que	leur parents euh parce que ça va pas dans leur famille quoi genre il y en
1129	1127	stag_bad_08_verifie_anonymise.trs	+ je crois que c'est des enfants qui sont placés euh parce que leur parents euh	parce que	ça va pas dans leur famille quoi genre il y en a leurs parents qui sont en t
1130	1128	stag_bad_08_verifie_anonymise.trs	et tout quand même les gosses jours elle me racontait elle s'est fait chier	parce que	jusqu'à quatre heures et demie les gamins ils sont pas là donc elle elle es
1131	1129	stag_bad_08_verifie_anonymise.trs	de neuf heures à seize heures bah pour elle c'est long pour les éduc- non	parce que	tu as tout parce qu'avec toute la paperasse qu'il y à faire maintenant euh
1132	1130	stag_bad_08_verifie_anonymise.trs	la paperasse qu'il y à faire maintenant euh tu vois eux du coup c'est bien	parce que	il y a pas les gamins donc du coup ils peuvent euh ils ont le temps de fair
1133	1131	stag_bad_08_verifie_anonymise.trs	+ et en fait à quatre heures et demie elle me disait en fait c'est super speed	parce que	à quatre heures et demie il faut aller chercher les gamins + en Trafic + da
1134	1132	stag_bad_08_verifie_anonymise.trs	é- l'E.S. elle était allée avec ¤NP tu vois pour euh lui montrer les écoles	parce que	il y a trois écoles différentes donc elle dit ouais je suis allée rechercher
1135	1133	stag_bad_08_verifie_anonymise.trs	ouais je pense ouais après c'est selon le cas et selon le décision du juge quoi	parce que	même quand ils font un placement le juge normalement est obligé de dé
1136	1134	tel_maz_07_verifie_anonymise.trs	peut-être même beaucoup des fois ça doit pas être facile ben pff oui et non	parce que	en général pas tous mais en général euh les enfants qui viennent à l'aide
1137	1135	tel_maz_07_verifie_anonymise.trs	d'enfants qui ont des problèmes de comportement ouais mais c'est ça le problème	parce que	lui niveau comp- comportement ça va encore enfin il est turbulent ave
1138	1136	tel_maz_07_verifie_anonymise.trs	sont euh plutôt euh + enfin on hésite pas à envoyer des enfants là-haut quoi	parce que	tu sais que euh elles sont relativement bien mais la plupart euh ça fait qu
1139	1137	tel_maz_07_verifie_anonymise.trs	sont relativement bien mais la plupart euh ça fait quand même on va dire peur	parce que	les gamins euh c'est des gamins avec des gros gros problèmes dé- de
1140	1138	tel_maz_07_verifie_anonymise.trs	ouais ouais ouais mais c'est vrai hein difficile que la vrai instit quoi	parce que	hum oh ça dépend parce que quand tu dis bon ils savent où ils vont don

Figure 2. *Fine Table generated automatically listing the occurrences of the entered keyword ("parce que"). The whole process takes less than 1 second for a list of 104 files and 1194 occurrences found.*

3. Instant data access

When the user clicks on any line of the excel table, a specific occurrence of the keyword is selected together with its left and right contexts, with span values of about 256 characters (rounded to the next word limit). The corresponding text and speech segments are then automatically retrieved and displayed, as shown in Fig 3 and Fig. 4.



☑ No case

☑ Exact

Left context: apprendre

Entry: parce que

Rigth context: j'étais

1 found

je vivais à part + alors mes enfants je pense pas qu'ils l'auraient acceptée cette vie-là non plus$ mais par contre mes enfants ils étaient en vacances en Tunisie$ c'est pas dans la famille mais ils ont été euh + avec des camps ils ça leur a plu +$ et je pense quand même que je retournerais en Tunisie avec mes enfants mes petits-enfants$

"L 1: tes enfants ils parlent l'arabe$

"L 2: ah non pas du tout ça n'a pas été facile pour leur apprendre l'arabe du fait que moi j'étais la seule$ mon mari français parlait pas l'arabe moi + parlant la + la langue française correctement$ je ne parlais pas l'arabe et ça m'arrive de parler un peu en arabe + avec mon frère quand on se trouve tout seuls$ mais + par respect pour mes enfants et mon mari +$ quand il y a mes on est tous autour de la table je ne parle jamais en arabe avec lui$ même si il se permet de me parler en arabe je réponds en français$

"L 1: ils ont jamais essayé d'apprendre l'arabe$

"L 2: ils ont essayé mais + le maître il est qu'ils ont eu +$ l'instituteur + qu'ils ont eu de tunisien il se prenait pour Dieu$ le Père du fait qu'en Tunisie on doit respecter l'instituteur$ ce qu'il dit ainsi de suite mes enfants ils n'ont pas pu supporter ça et puis ils ont refusé d'y aller et$

"L 1: un peu comme en France euh + au début du siècle ou euh$

"L 2: ben voilà justement + l'instituteur c'était quelque chose d'extraordinaire +$ alors le ce jeune-là il a gardé la mentalité-là$ mais il a oublié que c'est c'est des jeunes français qui voulaient apprendre l'arabe$ il aurait été un peu plus souple avec eux je j'en suis sûre ils auraient appris$ mais moi à la maison je ne pouvais pas leur apprendre parce que j'étais la seule qui parlait la langue-là +$ c'était pas facile avoir six enfants et essayer de leur apprendre l'arabe avec +$ je menais la barque à la maison il fallait tout faire leurs devoirs tout ça alors je pouvais pas$

"L 1: et ça te manque pas tu aurais pas aimé qu'un de tes fils ou de tes une de tes filles parle l'arabe$

"L 2: si j'aurais peut-être j'aurais aimé bon parler mais pfff$ je peux dire que je ne je n'y ai même pas pensé pas du tout +$ du fait qu'on se comprend on se comprend très très bien en français$ je vois pas où elle est l'utilité de se mettre à deux à parler$ une autre langue que les deux les autres sont autour de la table$

"L 1: ouais mais pour les traditions garder les traditions euh$

"L 2: ben les traditions je les ai pas aimées c'est pas pour les inculquer à mes enfants$

"L 1: ouais mais tu fais quand même euh certaines &lt; fêtes comme l'Aïd$

"L 2: oui l'Aïd &gt; tout ça parce que c'est dans le Coran +$ parce que ça c'est dans le Coran +$ l'Aïd les le l'Aïd El Kébir le la fête du sacrifice

Figure 3. *Automatic generation of text from alignment files and selection of the entered key word ("parce que"), highlighted with its immediate context.*

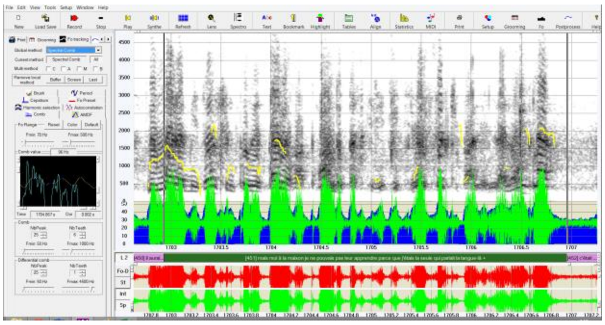


Figure 4. *Resulting display of the spectrogram, intensity and pitch curves corresponding to the segment automaticity retrieved from the Excel table.*

Integrating this function in one single software package makes possible specific research topics on prosody that would have been seen as too time consuming previously.

4. F0 foes

Whereas the integrated concordancer described above can be a valuable time saver, the actual acoustic analysis of prosodic events can prove disappointing when researchers are confronted to obviously erroneous fundamental frequency curves, while this parameter is one of the most important in prosodic analysis.

Indeed spontaneous speech recordings are often performed in noisy conditions, in places where echo and other not so obvious sources of problems are present (an example is given in Fig. 4). More specifically, the measurement of fundamental frequency is particularly sensitive to recorded speech signal distortions due to:

1) Poor signal to noise ratio;

2) Filtering of low frequencies eliminating low harmonics for male voices;

3) Harmonic blur due to room echo in the recording places;

4) Encoding in formats such as mp3 or wma with excessive compression levels;

5) Presence of external sound sources (car engine, overlapping speech segments, etc.);

6) Presence of creaky segments where the fundamental frequency is not really defined.

The speech analysis software Praat [7] for instance, de facto standard in this domain, revealed itself unsatisfactory for F0 tracking for a large number of recordings of the Rhapsodie project [10]. This leads first to evaluate the most frequent causes of F0 errors, then to elaborate various solutions in order to obtain reliable pitch curves. Among causes identified as sources of reliable speech pitch curves, we have:

1. Use of microphones with a poor response in low frequencies, resulting in the absence of the first harmonics in the spectrum (especially for male voices);

2. The presence of an important echo in the signal linked to the recording room dimensions, producing harmonic blurs. An unvoiced consonant can for example appear voiced due to the falsely observed continuity on the first harmonic;

3. A recording level too low, often due to an excessive distance between the microphone and the speaker, resulting in a low signal to noise ratio;

4. Use of AVC (automatic volume control) in the recording process, which distorts the speech intensity curve and indirectly producing errors the evaluation of vowel spectra;

5. Presence of multiple sound sources, in particular generated by low frequency engines (presence of a fridge in the recording room, etc.), or speech overlapping;

6. Excessive compression of the speech signal (e.g. wma or mp3 with a high compression parameter), giving when converted into waveform shifted spectral peak frequency

values undesirable for spectrum based algorithms (Cepstrum, Spectral Comb,…);

To address these potential problems en to ensure the generation of reliable F0 data, WinPitch has a catalog of methods applicable independently on user-selected speech segments:

Frequency domain methods

1. Spectral comb [4], obtained by correlation of the signal spectrum with a spectral comb with variable teeth intervals. Harmonics frequency range retained in the computation are user selectable;

2. Spectral brush [5], obtained by aligning signal harmonics on a selectable time window followed by a spectral comb analysis;

3. Cepstrum [9], evaluation of the periodicity of the log spectrum;

4. Swipep, developed by IRCAM, derived from the Swipe algorithm [2] based on harmonic detection followed by a Viterbi smoothing process;

5. Harmonic selection followed by spectral comb, with the retained harmonics selected by the user from a visual inspection on a simultaneously displayed narrow band spectrogram;

Time domain methods

6. Autocorrelation, operating directly on the speech waveform, available in three flavors, standard, normed Praat [1] and Yin [3], with adjustable window duration;

7. AMDF: average magnitude difference function, with the window length and the clipping percentage user adjustable;

8. Period analysis: F0 values are obtained from period's measurements from pitch markers placed automatically in a first pass and later manually corrected by the user;

These various methods give globally comparable results on good quality recordings. However, for lower quality recordings, the main problems can occur.

By nature, spectral based methods (such as the Spectral Comb) evaluate the signal fundamental frequency from the harmonic structure (i.e. the harmonic spectral lines of voiced segments), obtained from a Fourier transform. This requires an analysis signal time window relatively long (in the order of 32 ms or 64 ms for male voices with F0 equals to 100 Hz), which in turn prevents a correct tracking for fast rising or falling F0 values. The autocorrelation-based methods such as Yin may also exhibit this limitation even if they are based on the time domain (The reason stems from the time window usually selected for the autocorrelation). Other problems may arise when the fundamental frequency is very weak or absent (due to some filtering in the recording process for example);

The presence of pseudo-harmonics due to the presence of echo in the recording room can adversely affect frequency-based methods. The evaluation of the signal fundamental frequency of the Comb method for example is based on the detection of at least two consecutive harmonics. Echo produced by some harmonics, depending on the room dimensions, can generate trails of some harmonics long enough to make an unvoiced segment appear voiced (see Fig. 2) and confuse the algorithm detecting these components. It is quite difficult to differentiate automatically this spectral configuration from examples where a low frequency filtering would provoke spectral patterns similar to the ones generated by echo.

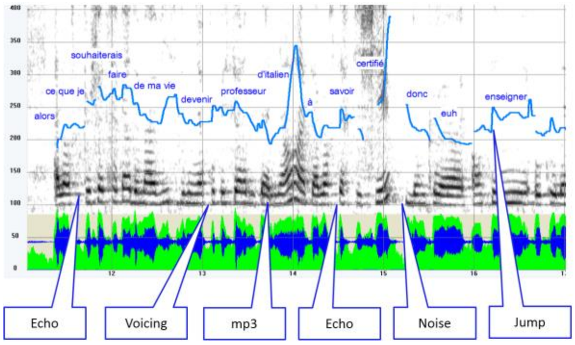
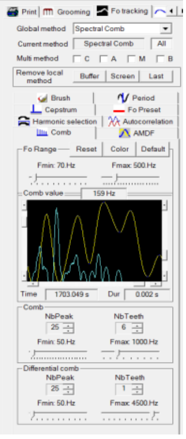


Figure 5. *Most common sources of errors for F0 tracking* (Rhap-D0003, PFC)

5. Cleaning F0 curves

To apply one of these methods, the user first selects a F0 tracking method in the command window (Fig. 6). Then a time window is selected on screen with the mouse guided by visual inspection of an underlying narrow band spectrogram. By releasing the mouse left button, the corresponding segment of the signal is automatically reanalyzed with the selected method, replacing F0 data with the new obtained values.

The new F0 curve segment is displayed in a color specific to the tracking method chosen, so that the user can identify visually on the overall F0 curve the tracking method pertaining to a specific time segment. Furthermore, by moving the cursor on screen, the corresponding command box corresponding to the F0 tracking method used for the wave segment defined by the cursor is displayed dynamically in the command box, together with all parameters values used for the chosen tracking method (Fig. 6).



Global method: Spectral Comb
Current method: Spectral Comb | All
Multi method: □ C □ A □ M □ B
Fmin: 70.Hz Fmax: 500.Hz
Comb value: 159 Hz
Time: 1703.049 s Dur: 0.002 s
Comb — NbPeak: 25 NbTeeth: 6 Fmin: 50.Hz Fmax: 1000.Hz
Differential comb — NbPeak: 25 NbTeeth: 1 Fmin: 50.Hz Fmax: 4500.Hz

Figure 6. Set of command boxes, for user selection of an alternate pitch-tracking algorithm applied locally on a speech segment.

A file containing all the information about corrections made can be saved in text format, as well as a .pitch file describing the corrected pitch curve to be exported to Praat.

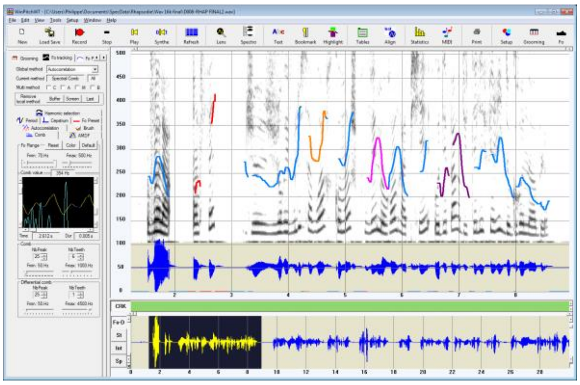


Figure 7. F0 curve sections are displayed in different colors according to the F0 tracking method used. The corresponding command box selected automatically on the left side (Rhap-D1001)

The two functions described above address the main concerns of researchers in the field of prosodic events analysis in their relationship with other structures of the sentence, syntactic and informational. The concordancer allows investigating a large number of occurrences of selected syntactic categories items, whereas the fundamental frequency "cleaning" gives reliable data in most cases retrieved by the concordancer, even in adverse recording conditions.

6. WinPitch as shareware

The software program is presently a shareware, whose installation code is free for the asking. WinPitch is downloadable from www.winpitch.com.

References

[1] Boersma, Paul (1993) Accurate short time analysis of the fundamental frequency and the harmonic-to-noise ratio of a sampled sound, Proc. Institute of Phonetic Sciences, 17. Univ. Amsterdam, 97-110.

[2] Camacho, Arturo (2007) Swipe: a sawtooth waveform inspired pitch estimator for speech and music, PhD thesis, University of Florida, 116 p.

[3] de Cheveigné, Alain and Hideki Kawahara (2002) Yin, a fundamental frequency estimator for speech and music. Journal of the Acoustical Society of America, 111(4).

[4] Martin, Ph. (1981) Extraction de la fréquence fondamentale par intercorrélation avec une fonction peigne, Proc. 12e Journées d'Etude sur la Parole, SFA, Montréal, 1981.

[5] Martin, Ph. (2008) Crosscorrelation of adjacent spectra enhances fundamental frequency estimation Proc. Interspeech, Brisbane, 22 – 26 September 2008.

[6] Martin, Ph. (2012) Automatic detection of voice creak, Proc. Speech Prosody, Shanghai, September 26-28.

[7] Praat, www.praat.org.

[8] Transcriber, a tool for segmenting, labeling and transcribing speech, http://trans.sourceforge.net/en/presentation.php

[9] Noll, A. Michael (1967) Cepstrum Pitch Determination, Journal of the Acoustical Society of America, Vol. 41, No. 2, (February 1967), 293-309.

[10] Rhapsodie (2010) Corpus prosodique de référence en français parlé, http://rhapsodie.risc.cnrs.fr/en/archives.html

[11] WinPitch, www.winpitch.com