Multi methods pitch tracking

Philippe Martin

CLILLAC-ARP, EA 3967, UFR Linguistique
Université Paris Diderot Sorbonne Paris Cité
philippe.martin@linguist.jussieu.fr

Abstract

The elaboration of rather large spontaneous speech corpora frequently implies the collection of data recorded with poor acoustic quality which may affect its acoustic analysis, and particularly fundamental frequency tracking (F0). Indeed, F0 analysis is particularly sensitive to distortion due to low signal to noise ratio, filtering of low frequencies, encoding in compressed formats (mp3, wma, …), room echo, not to mention the presence of external sound sources (car engine, overlapping speech segments, etc.).

In order to obtain a more reliable F0 analysis, it can be noted that some fundamental frequency algorithms are more reliable than others on specific voiced segments, depending on complex characteristics such as rate of F0 change, intensity of the first harmonic, presence of echo, etc.

For that reason a system (implemented in the software package WinPitch) is proposed to allow the user to select various tracking algorithms, adjust their parameters and apply a selected tracking method on the speech segments considered. The user is guided in this operation by an underlying narrow band spectrogram, which allows visual checking of the validity of the local F0 analysis by comparison between the F0 curve and the spectrogram low harmonics.

Index Terms: speech prosody, fundamental frequency tracking, intonation.

1. Introduction

Measurement of fundamental frequency is particularly sensitive to distortions such as low signal to noise ratio, filtering of low frequencies, encoding in compressed formats (mp3, wma, …), room echo, as well as the presence of external sound sources (car engine, overlapping speech segments, etc.). It is therefore important to ensure the calculation of reliable pitch curves even in adverse recording conditions, as these conditions may very well be associated to the most interesting examples from the linguistic point of view.

2. F0 foes

The use of the de facto standard speech analysis program Praat [9] to obtain reliable fundamental frequency curves in the Rhapsodie [1] project revealed to be unsatisfactory for a large number of recordings. In this project, about 2/3 of the files presented numerous problems for F0 analysis, among which:

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

b. The presence of an important echo in the signal, giving for example erroneous values of voicing in unvoiced stop segments;

c. 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;

d. Use of AVC (automatic volume control) in the recording process, corrupting the speech intensity curve and spectrum;

e. Presence of multiple sound sources, in particular generated by low frequency engines, or speech overlapping;

f. 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,…);

g. Use of an unnecessary high sampling frequency in the recording, involving large computation time and file size. This condition may affect F0 tracking due to unexpected rounding effects by the selected F0 algorithm.

3. Some F0 tracking methods

Since most pitch tracking algorithms are so far prone to errors in adverse recording conditions no matter their underlying operating principle, and given that for a particular speech segment some algorithms are less prone to errors than others, 8 different pitch tracking routines were implemented in the software program WinPitch [8] in order to evaluate fundamental frequency. These methods are spectral comb [5], spectral brush [6], autocorrelation (3 flavors: standard, Praat [2] and Yin [3]), AMDF, Cepstrum [4], spectral fit, and harmonic selection. Other methods (e.g. [7]) will be implemented in the near future.

These algorithms and their related parameters can be independently applied on user defined segments of the speech wave, in order to use the most satisfactory scheme for a given speech section of the recording.

WinPitch includes also a scanning feature allowing a quality analysis of the recording in terms of fundamental frequency coherence, transition and presence of creak (diplophonia and vocal fry varieties).

4. Visual detection of F0 problems

Most of the time, poor recording conditions affecting F0 tracking can be easily detected visually while displaying an underlying narrow band spectrogram. Although it requires some operator expertise and training, recurrent problems are usually presenting similar patterns for each category of problems.

The use of microphones with poor response in low frequencies or inappropriate low pass filtering in the recording chain is easily detected as long as higher frequency harmonics can be identified on the spectrogram. Spectral based pitch tracking algorithms usually perform better in these conditions (Spectral comb, brush…).

The presence of echo can be identified as long as it does not affect only the fundamental component in the spectrum (between two vowels it would then be confused with an actually missing fundamental). Fortunately in most cases, the recording room has dimensions large enough to generate echo of higher frequency. Large echo is particularly disturbing when affecting all harmonics of a rapidly rising or falling fundamental frequency. These identified echo segments must then be assigned a null F0 value.

The use of AVC gives generally a wrong detection of voicing, relatively easy to identify as generating incoherent F0 values at the end of voiced segments.

Undesirable sound sources present in recordings usually appear as relatively long segments of harmonics with constant frequency (engines or most musical sources). In the case of speech overlapping, harmonics evaluate differently on the time axis and can be separated, at least visually (except of course in the case of choir singing…). The spectral brush usually performs satisfactory in these conditions when the analyzed speech / second source intensity ratio is large enough.

Excessive compression of the recording speech gives recognizable fuzzy harmonic patterns on narrow band spectrograms. Usually time domain pitch tracking methods give better results in those cases.

5. Choice of F0 tracking methods

In order to improve the overall performance of the F0 analysis function, no less than eight fundamental frequency tracking methods are available to the user:

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

b. Autocorrelation in three flavors, standard, Praat [2] and Yin [3], with adjustable window duration;

c. Cepstrum [4];

d. Spectral comb [5], 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;

e. Spectral brush [6], obtained by aligning signal harmonics on a selectable time window followed by a spectral comb analysis;

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

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

h. Forced or imported value of signal section: on selected time intervals. The user can force F0 to be zero or be defined from imported values (from Praat for example).

To apply one of these methods, the user first selects a F0 tracking method in the command window (left of Fig. 1). 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 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. 2).

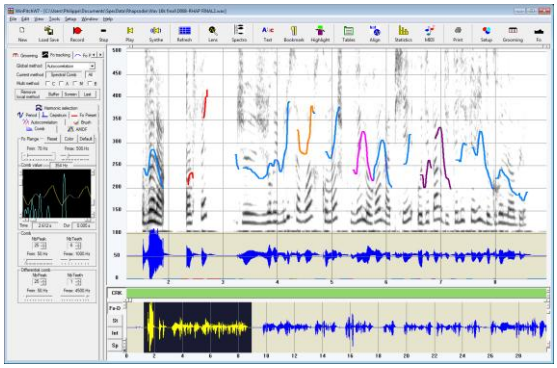

Figure 1. *F0 curve sections are displayed in different colors according to the F0 tracking method used.*

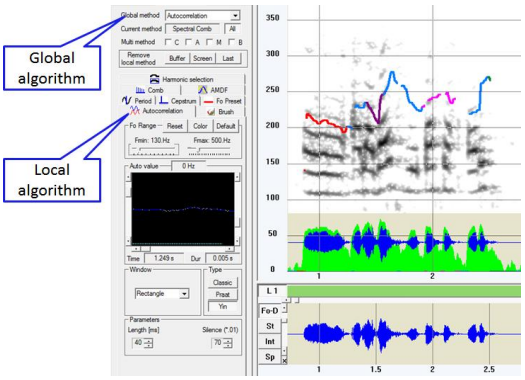

Figure 2. *Command box showing the selection of a global F0 tracking algorithm operating in a first pass on the signal, and a local algorithm acting on the user selected speech segment.*

6. Applying local F0 algorithms. An example

Fig. 3 shows an example of a difficult case for F0 tracking. The speech signal is recorded with low amplitude and the underlying narrow band spectrogram reveals a strong low pass filtering and the presence of noise and echo.

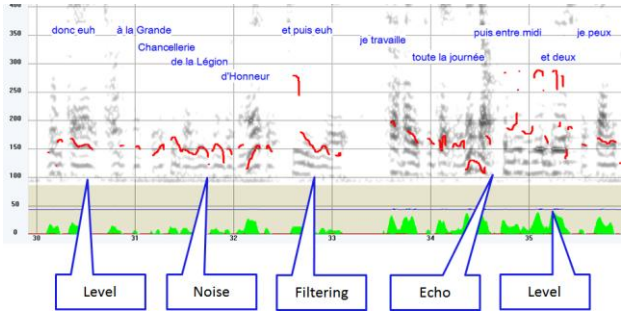

Figure 3. *A difficult example of F0 tracking on speech recorded with various distortions (low pass filtering, mp3 coding with low compression parameter, low amplitude level, echo …).Spectral comb is used as the global method.*

Fig. 4 to 7 display fundamental frequency curves obtained by various F0 tracking algorithms: autocorrelation, AMDF, spectral brush, harmonic selection, and the final F0 curve of Fig. 8 results from applications of various methods offered to the user on selected speech segments presenting problems, i.e. discrepancies between the F0 curve and low order harmonics displayed on the narrow band spectrogram.

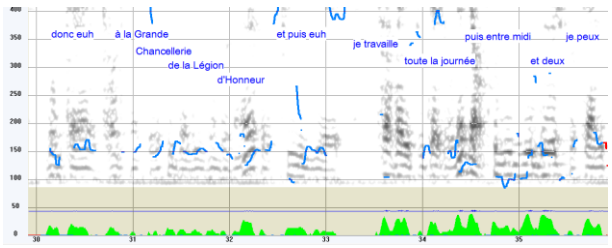

Figure 4. *Example of Fig. 3 using autocorrelation for F0 tracking.*

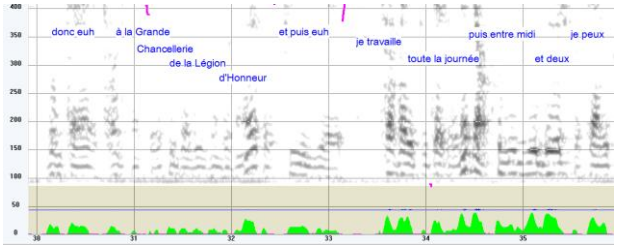

Figure 5. *Example of Fig. 3 using AMDF for F0 tracking.*

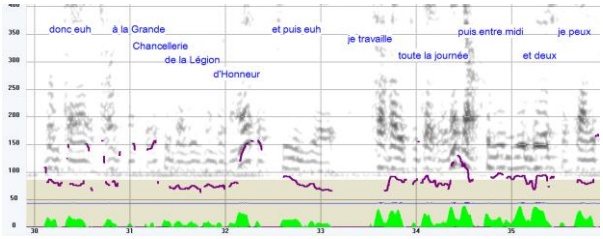

Figure 6. *Example of Fig. 3 using the spectral brush algorithm for F0 tracking.*

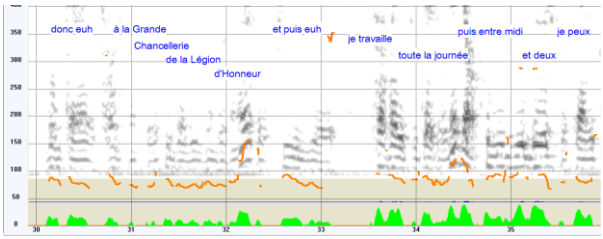

Figure 7. *Example of Fig. 3 using harmonic selection for F0 tracking.*

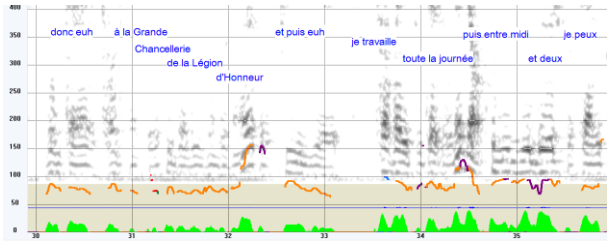

Figure 8. *Example of Fig. 3 showing the final F0 tracking cleaned locally using 5 different tracking methods.*

7. Applying an appropriate F0 tracking algorithm to a selected speech segment

An easy method to identify F0 errors is by comparison with the harmonics of an underlying narrow band speech spectrogram. Although this visual comparison does not allow for detection of all F0 errors (in particular for fast changing F0 values appearing as slow changing on a narrow band spectrogram due to the use of a large time window), it still permits a satisfactory localization of potential problems. Whereas jumps and doubling in F0 values are well known and

documented, effect of echo for instance is difficult to notice without visual inspection of low harmonics.

The case of echo, present in many examples of spontaneous speech recording made by inexperienced researchers, is particularly interesting since it can be confused with a voiced stop first harmonic. Only if the recording room presents such dimensions as to reveal an echo not affecting the fundamental frequency (for example the 3rd harmonic as shown in Fig. 9) can the user safely decide about the nature of the problem. If the echo is identified, the erroneous value can then be set to zero after selecting the corresponding speech segment.

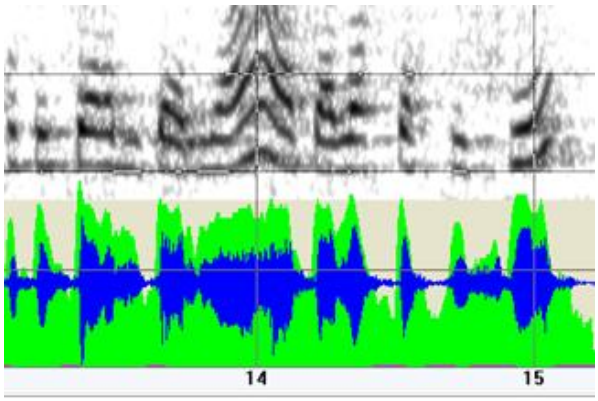

Figure 9. *An example of echo in the signal affecting the three first harmonics. A correct F0 curve can be obtained for this particular case by using the harmonic selection method.*

Once the identified problems have been corrected, all information entered by the user pertaining to the algorithm selected locally together with all relevant parameters is saved (in wp2 WinPitch proprietary format). When the recording is reloaded, all these parameters are restored in order to display the corrected F0 curve.

It is then possible to sample and save the F0 curve in various formats, including the .pitch format used in Praat.

8. Conclusion

All kinds of signal distortion can be regularly observed in spontaneous speech when recordings are made in normal life surroundings. As a reliable pitch curve is more than desirable in most circumstances, it becomes necessary to visually check the correctness of the F0 values, by comparing the pitch curves with the harmonics of an underlying narrow band spectrogram for instance. Once the problematic F0 curve segments have been identified, it becomes then most of the time possible to correct the pitch values of these segments by applying another F0 tracking algorithm that the one used by default.

Although this work intensive process can be rather tedious and requires a sound experience of pitch analysis from the operator, very good results can be obtained in what may be considered as desperate cases when fully automatic F0 tracking is used and fails. The implementation of this “pitch curve cleaning” process in WinPitch uses eight different F0 tracking methods and considerable care has been brought to the ergonomic aspects of the operations, in order to offer the users an efficient tool for reliable multi-method F0 tracking system.

Experiments are currently conducted to automatize partly or completely the F0 tracking selection process by analyzing specific acoustic conditions linked to recurrent errors in pitch values.

9. References

[1] Rhapsodie (2011) Corpus prosodique de référence en français parlé, http://rhapsodie.risc.cnrs.fr/en/archives.html
[2] 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.
[3] de Cheveigné, Alain and Hideki Kawahara. Yin, a fundamental frequency estimator for speech and music. Journal of the Acoustical Society of America, 111(4), 2002.
[4] Noll, A. Michael (1967) Cepstrum Pitch Determination, Journal of the Acoustical Society of America, Vol. 41, No. 2, (February 1967), 293-309.
[5] 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.
[6] Martin, Ph. (2008) Crosscorrelation of adjacent spectra enhances fundamental frequency estimation, *Proc. Interspeech, Brisbane*, September 26-28.
[7] Camacho, Arturo (2007) Swipe: a sawtooth waveform inspired pitch estimator for speech and music, PhD thesis, University of Florida, 116 p.
[8] WinPitch, www.winpitch.com
[9] Praat, www.praat.org.

Note: WinPitch can be downloaded from www.winpitch.com.