DOI: 10.17469/O2106AISV000026

PHILIPPE MARTIN

Prosodic Annotation of oral archives

Prosodic analysis of oral archives is often hampered by adverse recording conditions, making popular pitch tracking algorithm fail in many cases, so that gathered data may be useless. To address these limitations and allow prosodic research to be conducted on almost any kind of ancient or recent recordings showing less than optimal acoustic conditions, several dedicated functions have been integrated in the software program WinPitch. A first set of these functions allow the user to apply alternate pitch tracking algorithms among 7 available on selected signal segments, as some algorithms may perform better than others. If this process appears still unsatisfactory, a set of easy to use graphical commands are available to directly annotate pitch curves graphically, relying on an underlying narrow-band spectrographic display, whose frequency scale is automatically matched with the fundamental frequency curve scale. Piecewise linear graphic lines can be adjusted to melodic movements of any complexity as displayed on the spectrogram. Furthermore, the phonetic or phonological categories of these user placed annotations can be automatically labelled according to predefined classes, such as those available in ToBI notation system, or as melodic contours indicating dependency relations between stress groups.

Key words: Oral archives, prosodic annotation, sentence intonation, fundamental frequency.

1. *Introduction*

Oral archives found in old or not so old newsreels, radio and TV recordings are often characterized by less than optimal recording quality, making them unsuitable for prosodic research. In particular, the reduced speech bandwidth due to the then available technology for sound capture and conservation (carbon microphone, vinyl records, optical and magnetic tape sound recording, etc.) makes intonation analysis difficult and unreliable for some pitch tracking algorithms, such as the ones found in the popular software program Praat. Most analysis problems are linked to reduced signal intensity, lack of speech fundamental frequency in the spectrum, presence of echo and overlapping voices or sound sources, etc.

To address these limitations and pursue prosodic research on valuable speech archives, two sets of dedicated functions have been implemented in the software program WinPitch. Both sets rely on visual inspection and comparison of the pitch curve against the simultaneous displayed narrow-band spectrogram first (or second) harmonic, whose frequency scale has been aligned on the pitch curve scale. Using one of the seven tracking algorithms available in WinPitch (autocorrelation, AMDF, spectral comb, spectral brush, Cepstrum..., see Martin, 1981), the user is able to visually check the validity of a given pitch track section against the corre-

sponding spectrogram first harmonic segment, select graphically the F0 curve section of interest and apply an alternate algorithm on the selected section. As some algorithms perform better than others depending on the specific conditions of recorded segments, a more satisfactory F0 curve can be obtained, as possibly assessed by the underlying spectrogram first harmonic. Usually, switching between spectral comb and autocorrelation gives satisfactory results, the ancient AMDF method being surprisingly efficient when the other two algorithms fail.

When the situation is desperate, i.e. when no available pitch tracking method appears reliable enough, another set of user-friendly commands using only mouse controls allow to position graphic piecewise linear segments approximating any shape of the fundamental frequency curve. These graphic annotations can be user-labelled in 14 programmable categories (type, color, segment thickness, etc.) and once positioned on screen have their characteristics in time and frequency transferred to any spreadsheet program such as Excel in one single mouse click.

The user-defined classes can be adapted to virtually any phonetic or phonological model of sentence intonation, using for instance either the ToBI notation system or melodic contours notation.

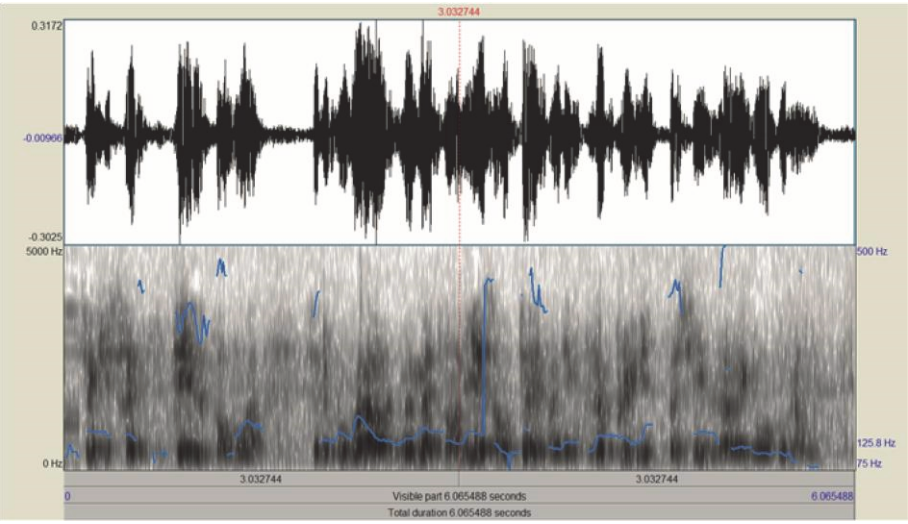
2. *Underlying spectrogram*

Although not frequently used, one of the best ways to validate the fundamental frequency curve of a given speech segment consists to compare the melodic curve with the first harmonic of the corresponding narrow-band spectrogram. Alternatively, the second or even the third harmonic can be used if the first is not visible, has been filtered or is obscured by another sound source spectral component. This is achieved in WinPitch with a single mouse click, aligning both spectrogram and fundamental frequency curve scales.

Except in rare case of a very fast changing fundamental frequency (as in in the transition from a voiced stop to a high intensity vowel), the spectrum first harmonic is totally reliable, although not allowing a precise measurement of the frequency value due to the frequency resolution of the spectrogram (in the order of 20 to 30 Hz).

In this domain of research, many users may not be aware that commonly available pitch tracking algorithms can display erroneous results and trust blindness the displayed fundamental frequency curves. Actually, a simple comparison with an underlying narrow-band spectrogram would give them some hints pertaining to the reliability of the acoustic analysis. Unfortunately, in many examples using Praat for instance (Fig. 1), when a spectrogram is displayed together with a pitch curve, it is by default a wide-band where harmonics and the fundamental component cannot be discriminated.

Figure 1 - *An example of pitch analysis of a noisy recording with Praat. Ce qui serait utile pour moi c'est de pouvoir me mettre en liaison téléphonique le plus vite possible avec mon gouvernement. "What would be useful for me is to be able to get in touch with my government as quickly as possible"*



3. *Automatic segmentation*

WinPitch integrates an automatic text to API segmentation and alignment system. Contrary to other realizations (e.g. EasyAlign, Maus...), the alignment algorithm proceeds by alignment with a reference generated by an embedded TTS system for each text segment. This approach integrates specific handling for liaisons and vowel linking specific for French. It allows the Viterbi based alignment to operate for more than 42 different languages (German, Italian, Spanish, European Portuguese, Swedish, Danish, Polish, Russian, Japanese, Korean, Mandarin...) based on the available TTS system in Microsoft Windows, so that there is no need to create new Gaussian models for each language as in EasyAlign. This is especially critical in the case of old speech recordings, where statistical approximation of phone properties is rarely valid. The automatic segmentation implemented in WinPitch generates automatically a segmentation layer words and API labelled phone units (Fig. 2).

Figures 3 to 5 illustrate the process applied to a recording made secretly on June 22nd 1940 by the German Wehrmacht in Rethondes. The French general Huntziger is negotiating the armistice while the German invasion of France is progressing. Fig. 3 shows the pitch segments manually placed along the spectrogram narrow-band first harmonic, and automatically labelled according to predefined criteria (terminal, major and minor continuation, neutralized contour).

Figure 3 displays the prosodic structure as indicated by the melodic contours of Fig. 2, and Figure 4 gives an example of the spreadsheet output corresponding to the example.

Figure 2 - *An example of prosodic annotation of a noisy recording. Segments of the melodic contours are graphically annotated to fit both the pitch curve and the corresponding spectrogram first harmonic. Ce qui serait utile pour moi c'est de pouvoir me mettre en liaison téléphonique le plus vite possible avec mon gouvernement. "What would be useful for me is to be able to get in touch with my government as quickly as possible"*

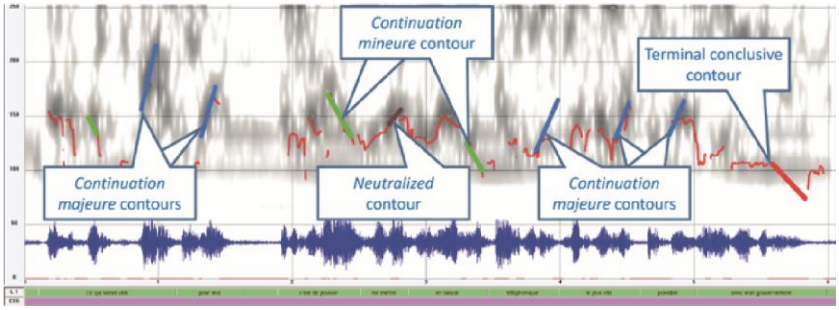


Figure 3 - *Automatic generation of the prosodic structure for the example of Fig. 3, illustrating the contrast of melodic slope characterizing French. Each contour indicated a dependency relation towards a contour situated later in the sentence and at a higher level in the Cneu-> Cfal-> Cris-> Cdec hierarchy*

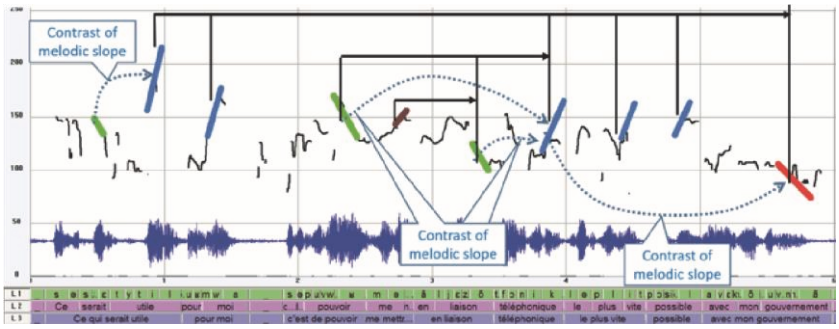


Figure 4 - *Output of the data on a spreadsheet giving the actual values the fundamental frequency, intensity and duration values of stress group segments values*

	File	Contour	Nb	Time [s]	F0 [Hz]	Int [dB]	Dur [ms]	Fmin [Hz]	Fmax [Hz]	Diff [Hz]	L 1
1											
2											
3											
4	Rothondes 1940	C2	1	0.475	149	34 S	69	134	149	15	Ce qui serait utile
5				0.544	134	23 S					
6											
7	Rothondes 1940	C1	2	0.873	156	23	109	156	215	59	Ce qui serait utile
8				0.982	215	35					
9											
10	Rothondes 1940	C1	3	1.322	132	35	103	132	177	45	pour moi
11				1.426	177	35					
12											
13	Rothondes 1940	C2	4	2.265	170	35 S	175	131	170	39	c'est de pouvoir
14				2.441	131	27 S					
15											
16	Rothondes 1940	Cn	5	2.723	143	27 S	86	143	156	13	me mettre
17				2.810	156	27 S					
18											
19	Rothondes 1940	C2	6	3.300	125	27 S	118	100	125	25	en liaison
20				3.418	100	28 S					

4. *Prosodic annotation*

Prosodic annotation is performed directly on screen, by matching the displayed curve segments with a piecewise linear curve placed on screen by the user with the mouse and automatically encoded in color and classified along the user predefined classes. These user-defined categories can be adapted to virtually any phonological model of sentence intonation, using for instance either the ToBI or melodic contours notation.

In the incremental prosodic structure model for example (Martin, 2018), sentence intonation uses melodic contours categories *Cdec*↓ (terminal falling and low for declarative sentences), *Cris*↗ (rising above the glissando threshold), *Cfal*↘ (falling above the glissando threshold) and Cneu→ (rising or falling below the glissando threshold. The corresponding phonological category is automatically assigned by the software to the segment placed by the user, together with the color coding eventually assigned beforehand. The glissando values (Rossi, 1971) are evaluated with the formula St2-St1/t2 – t1, with the threshold 0.32 (St2-St1) / (t2 – t1)^2, St1 and St2 are the beginning and end values of the contour in semitone units, and t2-t1 is the duration of the melodic contour.

Using the ToBI notation symbols such as H*H%, L*L-, L+H* for FToBI, it is equally easy to predefine as actual acoustic melodic movements are then approximated in terms of tone targets (Delais-Roussarie, Post, Avanzi, Buthke, Di Cristo, Feldhausen, Jun, Martin, Meisenburg, Rialland, Sichel-Bazin & Yoo, 2015). Targets can be acoustically predefined, with values of fundamental frequency jumps (in semitones) for example, and assigned automatically or manually.

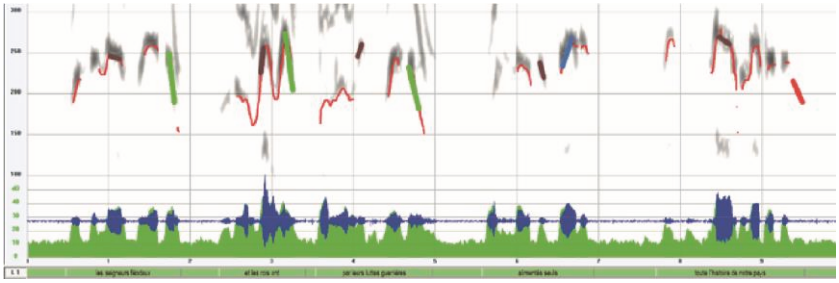
5. *Some oral archives*

The described integrated system has been applied to the analysis of some number of speech archives, both old and recent. Examples pertaining to the changes occurred in the realization of melodic contours in French speech archives (early XX century period) are given below, showing the capability of the system in degraded recording conditions to illustrate the evolution of prosodic style in political and news speeches of these periods.

Available speech archives recorded before 1918 are essentially characterized by the use of tremolo segments by the speakers, but apparently only used in very formal circumstances. Such tremolos realizations are not found in the same period in non-formal situations, where the contrasts of melodic slope, found as well in contemporary recordings, was already implemented in the phonological system of French.

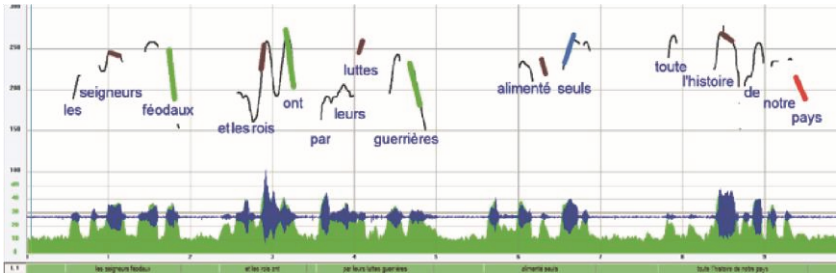
A first example pertains to Henry Le Châtelier (1850-1936), recorded in 1910 (Fig 5 and 6). *Les seigneurs féodaux et les rois ont par leurs luttes guerrières alimenté seuls toute l'histoire de notre pays* "The feudal lords and kings have by their war fights fueled alone the entire history of our country".

Figure 5 - *Recording of H. Le Châtelier in 1910: Les seigneurs féodaux et les rois ont par leurs luttes guerrières alimenté seuls toute l'histoire de notre pays. "The feudal lords and kings have by their war fights fueled alone the entire history of our country"*



Since the fundamental frequency is almost totally absent from the narrow-band spectrogram, it is the second harmonic that makes it possible to validate the plot of the melodic curve obtained by the spectral comb method.

Figure 6 - *The recording of Fig. 5 aligned with text, showing the contrast of melodic slope and the absence of tremolo found in a more formal speech style*



The annotated melodic contours show the incremental merging of successive stress groups forming the overall sentence prosodic structure: [[*Les seigneurs Cneu→ féodaux Cfal*↘] [*et les rois Cn→ ont Cfal*↘] [*par leurs luttes Cn→ guerrières Cfal*↘] [*alimenté Cneu→ seuls Cris*↗]] [*toute l'histoire Cneu→ de notre pays Cdec*↓]

The melodic slope contrasts are constantly realized by the speaker, with an unusual segmentation in the group *et les rois ont*, and there is no trace of tremolo as found in speech formal recordings of the same period as shown below.

By contrast, the analysis of a speech given by Fernand Brunot in 1911 shows a large use of pitch tremolo on stressed vowels, where the phonological system of French would expect a *continuation majeure*, instantiated by a rising melodic contour above the glissando threshold (Fig. 7 and Fig. 8). Ferdinand Brunot was professor of linguistics at the Sorbonne University in Paris from 1900- to 1934. He became very interested in the available speech recording techniques of his time, and especially to speech pronounced by ordinary people. The segments analyzed in Fig. 8 and Fig. 9 come from his inauguration speech given on June 3rd, 1911 for the inauguration of the *Archives de la parole* he created with Émile Pathé.

Figure 7 - *Ferdinand Brunot in 1911... de tant de beauté de gloire et d'espérance de tant d'accord si doux d'un instrument divin... "so much beauty of glory and hope of so much sweet agreement of a divine instrument..."*

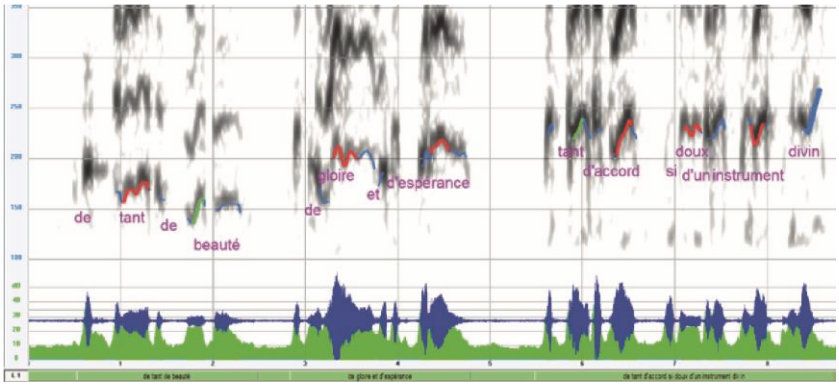
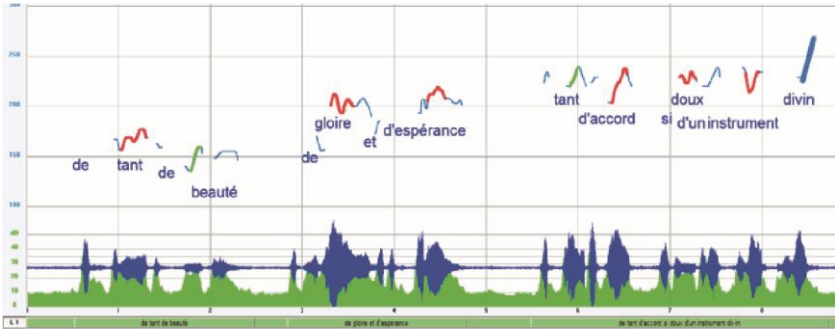


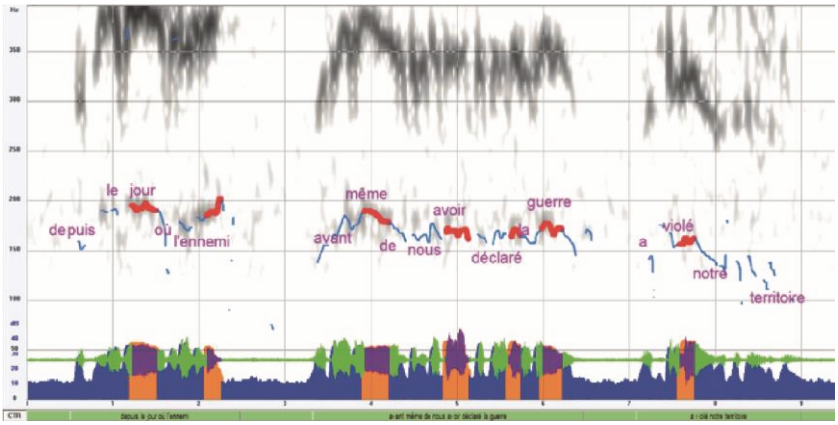
Figure 8 - *Ferdinand Brunot in 1911... de tant de beauté de gloire et d'espérance de tant d'accord si doux d'un instrument divin... showing the large use of tremolos on stressed vowels*



Rising emphatic stress on the first syllable of *beauté* as well as the first syllable of the group *tant d'accord*. Despite the presence of many syllables realized with tremolos, Brunot sill produces a first *Cris* contour on the last syllable of the sentence.

The following example of formal speech (Fig. 9) also illustrate the use of tremolos. The speaker is Paul Deschanel then president of the *Assemblée nationale* at the time of the recording, just before the beginning of the first World War.

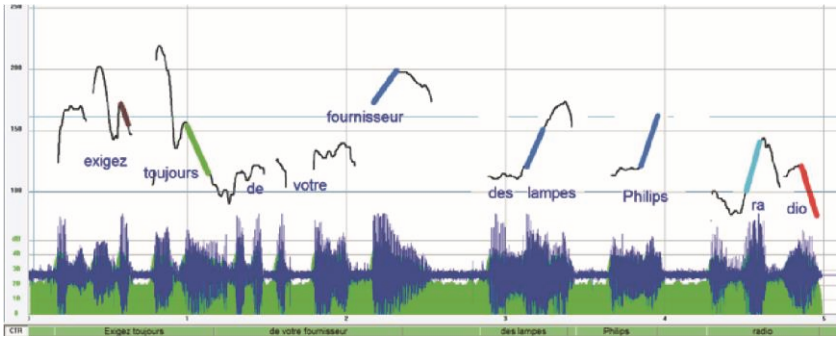
Figure 9 - *Paul Deschanel 1914. Depuis le jour où l'ennemi avant même de nous avoir déclaré la guerre a violé notre territoire. "Since the day when the enemy even before we declared war has violated our territory"*



Without tremolos, the realization with melodic contours would have been: *Depuis le jour Cfal*↘ *où l'ennemi Cris*↗ *avant même Cfal*↘ *de nous avoir Cneu*→ *déclaré Cneu*→ *la guerre Cris*↗ *a violé Cneu*→ *notre territoire Cdec*↓

Fig. 10 gives an example of radio advertising recorded in 1928, where the contrast of melodic lope is clearly illustrated.

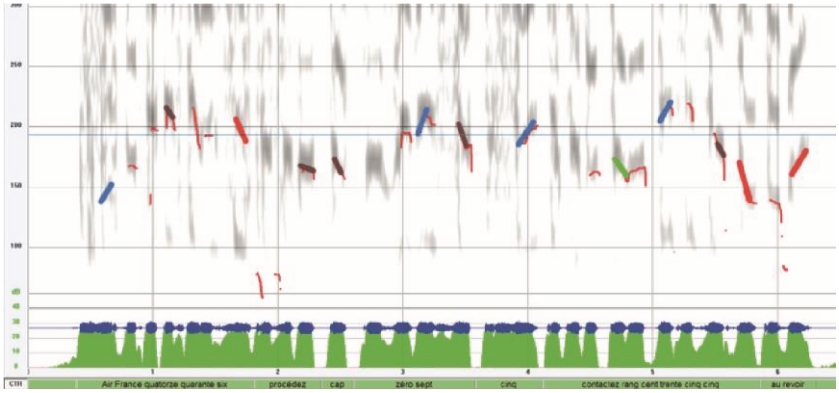
Figure 10 - *Radio advertising in 1918. Exigez toujours de votre fournisseur des lampes Philips radio. "Always ask your supplier for Philips radio lamps"*



The tone of voice is clearly non-formal in this example. The sequence of melodic contours on stressed vowels shows a contrast of melodic slope on *Exigez toujours Cfal*↘ *de votre fournisseur Cris*↗, whereas the first syllable of radio bears an emphatic stress (always rising) on the first syllable.

Figure 11 provides a last example, an air traffic control recorded in 1996, while more recent, shows adverse recording conditions for prosodic research. In this example, the fundamental frequency is almost totally absent in the signal, due to a severe filtering of components below 300 Hz (as in old telephone land lines).

Figure 11 - *An air traffic control recorded in 1996. Air France Cris ↗ 14 Cneu →46 Cdec↓ procédez Cneu→ cap zéro Cris ↗ sept cinq Cris ↗ contactez rang cent trente cinq cinq Cdec↓ au revoir ↗. "Air France 1446 proceed cap zero seven five contact rank one hundred thirty five five goodbye"*



6. *Conclusions*

The oldest speech recordings available in French indicate that, if some phonetic realization of melodic contours did change in some 100 years (e.g. no more penultimate stress, ex.: on suffixes *-ation*, rare tremolos after 1918 in formal speech), the phonological contrast of melodic slope, a feature specific to French, did not change and was already effective at this time.

The prosodic annotation function implemented in WinPitch allow for the analysis of desperate cases, where no fundamental frequency tracking algorithm seems to be suitable.

Although it requires some basic expertise from the annotator, the semiautomatic annotation constitute a valuable approach for the study of old and degraded speech archives. WinPitch is available at www.winpitch.com.

Bibliography

Delais-Roussaire, E., Post, B., Avanzi, M., Buthke, C., Di Cristo, A., Feldhausen, I., Jun, S-A., Martin, Ph., Meisenburg, T., Rialland, A., Sichel-Bazin, A. & Yoo, H-Y. (2015). Intonational phonology of French: Developing a ToBI system for French. In Frota, S., Prieto, P. (Eds.), *Intonation in Romance*. Oxford: Oxford University Press, 63-100.

EASYALIGN. http://latlntic.unige.ch/phonetique/easyalign.php.

Martin, Ph. (1981). Extraction de la fréquence fondamentale par intercorrélation avec une fonction peigne. In *Proc. 12e Journées d'Etude sur la Parole, Montréal, 1981*, 221-232.

Martin, Ph. (2018). *Intonation, structure prosodique et ondes cérébrales.* London: ISTE.

MAUS. https://www.bas.uni-muenchen.de/Bas/BasMAUS.html.

PRAAT (2019). Computer program, www.praat.org.

Rossi, M. (1971). *Le seuil de glissando ou seuil de perception des variations tonales pour la parole.* In *Phonetica*, 23, 1-33.

WINPITCH (2019). Computer program, www.winpitch.com.