Automatic extraction of the general tendency of F0 patterns Cross linguistic study on laboratory data

Katarina Bartkova, Mathilde Dargnat

University of Lorraine, ATILF, France
katarina.bartkova@atilf.fr, mathilde.dargnat@atilf.fr

Abstract

The goal of our study is to use an automatic approach to extract the general prosodic tendencies of the speech signal conveyed by the F0 pattern. The speech signal is prosodically annotated by an automatic prosodic transcriber and then prosodic patterns are extracted from this annotation. The pertinence of the pattern extraction is tested here on laboratory data containing isolated sentences in French and English uttered by native and non-native speakers. An analysis of the extracted parameters shows how the prosody of the sentences is defined by their shared syntactic structures and also indicates to what extent the prosodic features used by the two languages are similar or different. It appears from the analyzed data that the extraction of parameters via automatic processing can yield relevant information for a cross-linguistic study of prosody.

Index Terms: prosodic annotation, automatic pattern extraction, native & non-native prosody

1. Introduction

The use of an automatic approach for prosodic annotation of speech is useful, especially as agreement on manually annotated prosodic events (boundary levels, disfluences and hesitation, perceptual prominences) between expert annotators is quite low [15]. When manual coding of pitch level is carried out, there is the risk that human annotators can be influenced by the meaning of the speech. Moreover a human transcriber may be also influenced by what he considers to be the norm, thereby standardizing the transcription of prosodic phenomena and ignoring the reality of the speech signal.

A further advantage of automatic processing is that, once the values of the parameters are normalized, they are then compared to the same threshold value. This is difficult to achieve with manual annotation because of the inherent subjectivity of this approach.

The goal of the present study is to extract relevant prosodic tendencies of the F0 pattern. This approach is then tested in a cross-linguistic study of speech prosody in French and English.

2. French & English prosody

Many studies have described the specificities of French and English prosody. According to these studies, French uses a combination of segmental and tonal cues to signal prosodic phrases, and differs in this respect from a language like English, which relies almost exclusively on tonal boundaries [7] [14]. In French, lexical stress is mostly quantitative [8], and the final syllable is the one which undergoes a potential lengthening. However, lengthening of the last syllable of the word corresponds also in French to final (pre-boundary) lengthening, which concerns rhythm, and is not an accentual lengthening as in English [6

French is generally considered as a language with mostly 'rising' F0 patterns [12] accompanied by a lengthening of final syllables [20]. French prosodic phrasing was described by Delattre's functionalist approach [9] Though extended by more recent studies [11], [13], [10]. Delattre's work still remains seminal for studies on French prosody. In French spontaneous speech data, a melodic rise is generally produced at the end of the clause. It indicates that the clause is an unfinished constituent at the discourse level, and that it can be associated with the term of "**major**" or "**minor**" **continuation contour**, according to Delattre's approach.

French and English intonation are sometimes described by a set of contours. Delattre [9] considers that 10 basic contours can describe the most frequent intonation patterns in French; [18] also distinguish 10 contours though their contours differ from those proposed by Delattre. As far as English is concerned, 22 pertinent intonation contours are proposed by [17] to describe English intonation.

It is common to use the term of *assertion intonation* or *question intonation* to refer to falling or rising contours: falling contours are associated with assertion or assertiveness (Bartels 1999), whereas rising contours are associated with questions or aspects of questioning (uncertainty, ignorance, call for a response or feedback from the addressee, etc.). Although prototypical assertions are uttered with a falling contour and prototypical confirmation or verifying questions are uttered with a rising contour, occurrences of assertions with a rising contour and occurrences of confirmation or verifying questions with a falling contour are far from rare in everyday conversations [4].

In the following paragraphs F0 contours in French and English sentences spoken by native and non-native speakers are measured and compared and their differences are statistically evaluated.

3. Prosodic annotation

Prosodic parameters are subject to parameter values governing prosodic coherence along the prosodic group. It was observed in automatic speech processing (in diphone and data driven speech synthesis) that a sudden unjustified change in F0 or sound duration (beyond stressed syllables or prosodic junctures), is perceived either as a corruption of the speech signal or as an occurrence of a misplaced contrastive stress [5]. Most of the time researchers focus on the transcription of parameter values of syllables considered as linguistically prominent, carrying pertinent linguistic information. The other linguistically non prominent syllables, remain generally uncoded although their prosody contributes to an overall perception of a correct pattern. Therefore we believe that in order to keep a faithful prosodic transcription of the speech signal, all the parameters of the syllables should be annotated.

3.1. Prosodic labelling

Speech data processing was carried out in several stages. First, prosodic parameters were extracted from the speech signal. In order to segment the speech data, a text-to-speech forced alignment was carried out using the CMU sphinx speech recognition toolkit [16] yielding an automatic segmentation of the speech signal at the phoneme level. This automatic segmentation of the speech signal was then manually checked by an expert phonetician.

For the F0 pattern analysis, F0 values in semitones were estimated every 10 ms by the software Aurora [19]. A simple F0 parameter smoothing was carried out by our annotation software to eliminate corrupted F0 values.

Prosodic annotations were yielded by the language independent automatic annotation tool PROSOTRAN [2]. This tool requires no specific linguistic knowledge, therefore it is well-adapted for cross-linguistic studies. PROSOTRAN yields various numeric and symbolic prosodic annotations for each syllable of the speech signal; however, from this data, only F0 range values and sound durations are used in this study. Sound duration is normalized and transformed to a symbolic duration annotation. For the representation of F0 patterns, a melodic range is calculated between the maximum and the minimum values of the F0 in semi-tones. All speech material for each speaker is used to build a histogram of the distribution of the F0 values. To avoid extreme, often wrongly detected F0 values, 6% of the extreme F0 values (3% of the highest and 3% of the lowest ones) are discarded. The resulting range is then divided into several zones (9 in our case) and is coded into levels (from 0 to 9). By calculating FO in this way, value normalization is enabled and also inter-speaker comparison of FO patterns.

3.2. Corpus

The corpus used in this study was recorded as part of the Intonal project, focusing on the study of intonation in French and English. The recorded corpus contains 40 short sentences belonging to 8 syntactic categories using 20 French and 20 English native speakers. The French speakers uttered French and English sentences, and constitute our non-native English speaker group.

The corpus sentences contain sentences with two kinds of non-conclusive F0 slope configurations as well as interrogative and declarative sentence final configurations. Our study analyses mainly the F0 contours on discourse level (the F0 value of the final segment of declarative clauses connected by a discourse relation, marked or not by a conjunction) and on syntactic level (F0 pattern on the final segment of declarative and interrogative sentences).

3.2.1.1 F0 tendency extraction

The goal of the F0 pattern extraction is to get the most representative F0 pattern(s) for a given sentence for a group of speakers keeping one or, if necessary, several F0 values per syllable (Figure 1).

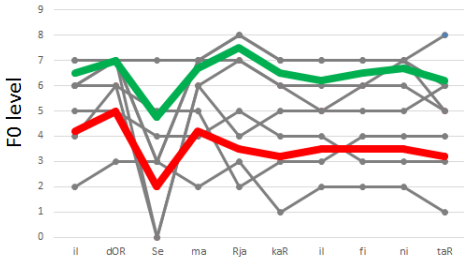


Figure 1: *Representative patterns (in green and red) of F0 levels for a sentence uttered by 8 speakers (in grey)*

The symbolic annotation yielded by PROSOTRAN is used for each syllable and for the whole speaker group of each language to identify the values that represent the general tendency of the F0 pattern. An empirical approach was adopted allowing the emergence of maximally two F0 values per syllable (as the number of speakers is relatively small in each group – maximally 18 for English native speakers). The F0 values coded by their range level are split into two groups using an adjusted median value keeping F0 values belonging to the same symbolic code in the same group. This way the division of the following symbolic F0 values [10 9 9 8 8 8 7 7 5 6] occurs after the last "8" (the 6th F0 value and not after the 5th value as expected). Each grouping obtained is represented by a mean value (V1 & V2, cf. Figure 2) and the two F0 values per syllable are maintained only if their difference is higher than 3 semi tones and when the number of F0 values in a group is higher than 2. If not, the groups are merged and a general mean value (V(1,2)) is calculated using the values of all the speakers for a given syllable.

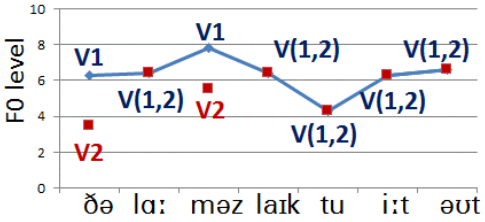


Figure 2: *F0 pattern general tendency: V1 first F0 value, V2 second F0 value, V(1,2) first and second F0 values merged*

The number of merged values was 70% for the French native speakers, 74% for English natives but only 50% for non-native speakers. The F0 patterns of non-native speakers were less consistent and have more variability in their pronunciation.

For the different sentences, the succession of the F0 values is recovered and the preferred F0 pattern tendency observed. For example, for the sentence in figure 2, the preferred tendency as to the succession of the F0 values is represented on Figure 3.

6.3 (V1) 6.4 (V1) 7.8 (V1) 6.4 (V1) 4.3 (V1) 6.3 (V1) 6.6 (V1)	**[4]**
3.5 (V2) 6.4 (V1) 5.5 (V2) 6.4 (V1) 4.3 (V1) 6.3 (V1) 6.6 (V1)	**[2]**
6.3 (V1) 6.4 (V1) 5.5 (V2) 6.4 (V1) 4.3 (V1) 6.3 (V1) 6.6 (V1)	**[2]**

Figure 3: Succession of the F0 range values and their codes (between parentheses) and the number of the pattern observed (between brackets)

3.3. Analysis of results

Our approach of automatic extraction of F0 tendency is tested on 4 sentence types from our laboratory data; that is on continuative, paratactic, interrogative and declarative sentences. Each sentence group contains 5 different sentences

of the same syntactic structure uttered by a group of at least 8 speakers - that is a corpus of about 160 sentences. In the following paragraphs only results obtained for one sentence (containing about 8 utterances) per sentence type will be discussed, however the results obtained for the remaining sentences of the same sentence type obtained very similar results.

Continuative sentences: *(two clause sentence, with coordinating conjunction, ("He has seen **Maria** because he has come" "Il va voir **Maria** car il en a envie").* (cf. Figure 4).

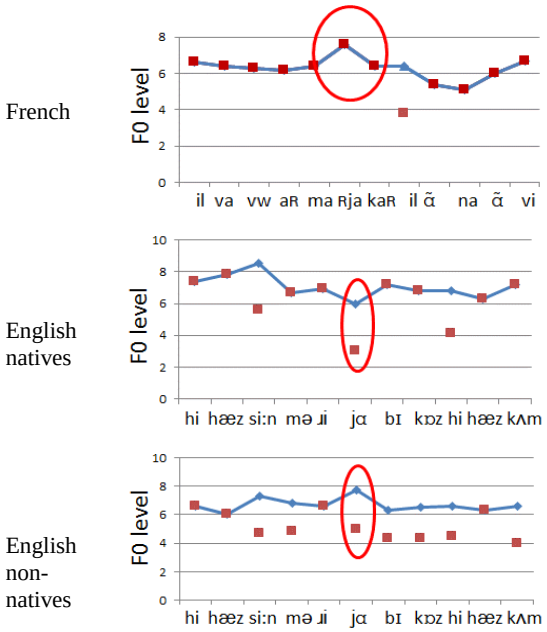


Figure 4: *F0 pattern in a continuative sentence ("He has seen **Maria** because he has come" "Il va voir **Maria** car il en a envie"); red circle: major prosodic boundary*

French speakers marked the continuation (red circle on the figure) with a rising F0 while English speakers prosodically coded the same syntactic boundary with a lowering F0. Non-native English speakers use more rising patterns than falling ones. In French, the general rising tendency of the F0 is not very high but the prosodic boundary is also indicated with lengthened vowel duration. On the other hand, the downwards movement of the F0 in English was more important but there is no vowel lengthening in the final syllable. A high prosodic agreement in these sentences is in the realization of the major prosodic boundaries: the rising tendency on the NP boundaries is respected by the majority of the French speakers and the falling F0 pattern by the majority of English speakers. In the non-native group there is little agreement as to the F0 pattern on major prosodic boundaries; in fact most of the time two F0 values are extracted: a high value indicating a rising F0 movement and a low value indicating a falling F0 movement.

Paratactic sentences: *(two clause sentence, without coordinating conjunction, "Il dort chez **Maria**, il va finir tard. / He'll sleep at **Maria's**, he'll finish late.")* (cf. Figure 5)

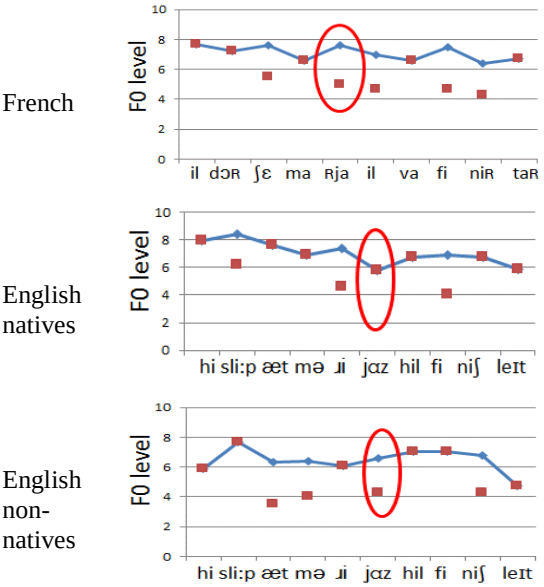
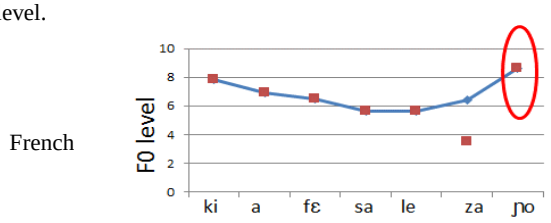


Figure 5: *F0 pattern in a paratactic sentences ("He slips at **Maria's** he'll finish late" "Il dort chez **Maria**, il va finir tard"); red circle: major prosodic boundary*

In French little prosodic agreement is found on the major non-final prosodic boundary (red circle on the figure): the F0 level fluctuated and most of the time two F0 level values are extracted. In these sentences the speaker's prosodic production of the first clause is similar to a final prosody signaling the end of the first clause in a sentence and its syntactic independence from the second clause. However, a strong agreement in the F0 values is observed in English native data where all the clause final F0 values are falling. As far as the non-native English group is concerned, their F0 realizations are again closer to the French group than to the English group.

Interrogative configuration *(simple subject NP: "Qui a fait ça? **Les agneaux**? / Who did this? **The lambs**?")* (Figure 6)

In French interrogative sentences, generally, a huge level rise is preceded by a rather flat F0 level. The F0 pattern in English interrogative sentences contains a more moderate F0 upward movement or (Figure 6) a lowering F0 movement (despite the interrogative character of the sentence). As the interrogative character of the sentence is also expressed by question words ("who" in this example), there is probably no real need for prosodic marking. Again, the French speakers of English are closer with their F0 pattern realizations to French prosody, (rising F0 values) than to English prosody. Other noteworthy findings for this sentence type: the interrogative character is prepared from the beginning (onset) of the sentence: for the 3 speaker groups the interrogative sentences start at a high F0 level.



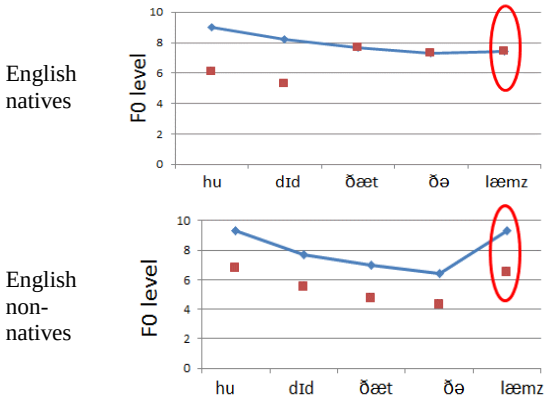


Figure 6: *Interrogative sentence types ("**The lamas** like to eat oats" "**Les lamas** aiment bien l'avoine"; red circle: interrogative sentence final boundary*

Declarative sentences *(Longer declarative sentence: "Il dort chez **Maria**. /He'll sleep at **Maria's**")* (cf. Figure 7)

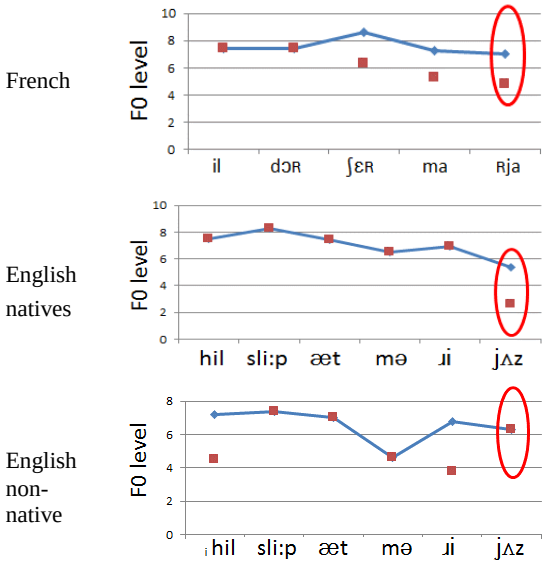


Figure 7: *F0 pattern in a sentence with continuative configuration on a subject NP ("**The lamas** like to eat oats" "**Les lamas** aiment bien l'avoine"); red circle: final declarative boundary*

In French the preferred F0 pattern at the end of the declarative sentence types is only slightly falling or it remains flat. This finding can be partly explained by the fact that, in French, the final pattern of a declarative sentence is also marked by the last syllable duration lengthening which is more moderate than the syllable duration lengthening on major continuation prosodic boundaries [1]. In English, on the other hand, the end of the declarative sentences is marked by a systematically falling F0 pattern. As for the non-native speaker group, their F0 pattern is also either slightly falling or remains flat i.e. they are closer to the native French group F0 patterns than to the native English group patterns.

3.4. General discussion

From the previous analysis some general tendencies can be identified as to the differences between French and English F0 patterns. In the phrases studied here French speakers gave preference to more rising F0 patterns on prosodic boundaries while English native speakers uttered the English version of the sentences with a preferential falling F0 pattern. The final F0 movement in the sentence is falling in both languages (cf. Table I), however the slope of the F0 level change is steeper in English than in French.

The English non-native speaker's prosody remains somehow influenced by both French and English prosody: for example in interrogative sentences their F0 pattern is clearly similar to French, while in continuative sentences the non-native's F0 pattern is more similar to English (cf. Table I).

Table I: *Amount of F0 levels in rising (+) or falling (-) patterns on major prosodic boundaries for the 4 sentence types*

Sentence type	French	English native	English non-native
Continuative	+2.6	-1.5	-1.5
Paratactic	+1.1	-1.5	-0.5
Interrogative	+3	+0.5	+2.5
Declarative	-1.8	-3	-1.5

The intra-speaker variability of the F0 levels (2 F0 level values per syllable) of the native (French & English) groups occurs more often in linguistically less important syllables. So in French only 1.7 syllables/sentence type (or 35% of linguistically pertinent syllables), and in English 1.5 syllables/sentence type (or 30% of linguistically pertinent syllables) are coded by two F0 levels on the linguistically pertinent prosodic boundaries. On the other hand, non-native speakers are less consistent in their prosodic production with more variability in F0 values: 2.8 syllables/sentence type (or 55% of linguistically pertinent syllables) are captured by 2 F0 level values.

Finally, variability is observed also with respect to the number of F0 levels used by the speakers of the 3 groups: the French and English native speakers used up to 6 F0 levels for their F0 patterns while the non-native speakers used only 4 F0 levels. That means that the F0 patterns of non-native speakers are more monotonous than the F0 patterns of the native speakers.

4. Conclusions

The goal of our study was to use an appropriate coding schema for prosody representation in a cross linguistic study of French and English prosody. The data used for testing the proposed method were laboratory data produced by a group of French and English native speakers and they contained sentences sharing the same syntactic structures in both languages. This syntactic specificity of the data base was well adapted to a cross-linguistic study as it allowed for comparison of prosodic phenomena relatively easily.

The methodological problem addressed here was how to represent prosodic parameters in such a way that comparison of the occurrences of these parameters in different sentences and languages would be pertinent. The aim of the study was to represent the general tendency of the F0 pattern by extracting one, or maximum two, F0 values per syllable, coded in terms of 9 F0 levels calculated from the voice range of each speaker. In the present study for each syllable one or maximum two F0 values were kept to capture the prosodic tendency of the sentence. However, in future, a more general automatic decision algorithm should be used to make the decision of the number of representative values of the F0 more data driven.

5. References

[1] Bartkova, K., Sorin, C.: A model of segmental duration for speech synthesis in French. Speech Communication 6(3): 245-260, 1986.

[2] Bartkova, K., Delais-Roussarie, E., Santiago-Vargas, F.: "PROSOTRAN : a tool to annotate prosodically non-standard data", Proceedings of Speech Prosody, Shangaï, China, 22-25 mai, 2012.

[3] Bartels, C.: *The Intonation of English Statements and Questions*, New-York: Garland Publishing, 1999.

[4] Beyssade, C., Marandin, J.-M., Rialland, A.: "Ground / Focus: a perspective from French". In R. Nunez-Cedeno *et al. (*eds), *A Romance perspective on language knowledge and use: selected papers of LSRL 2001*. Amsterdam/Philadelphia: Benjamins. pp. 83-98, 2003.

[5] Boidin, C. : *Modélisation statistique de l'intonation de la parole expressive*, Thesis, Université Rennes 1, 2009.

[6] Campbell, W.N. : "Syllable-based segmental duration". In *Talking machines: theories, models and design*, Bailly & Benoît (eds). Amsterdam: Elsevier, 211-224, 1992.

[7] Crystal, D.: *Prosodic systems and intonation in English*, Cambridge University Press, 1969.

[8] Delattre, P.: "A comparative study of declarative intonation in American English and Spanish", *Hispania* XLV/2, pp. 233-241, 1938.

[9] Delattre, P. : "Les dix intonations de base du français", *The French Review*, 40/1, 1-14, 1966.

[10] Delais-Roussarie, E. : "Vers une nouvelle approche de la structure prosodique", *Langue Française*, 126 : 92-112. Paris: Larousse, 2000.

[11] Di Cristo, A. : A propos des intonations de base du français. Unpublished ms., 2010.

[12] Fónagy, I., Bérard, E. : "Questions totales simples et implicatives en français parisien, Interrogation et Intonation: Studia Phonetica no 8, Ed. by Grundstrom A., Léon P. Paris: Didier. pp. 53-98, 1973.

[13] Fónagy, I. : "L'accent français, accent probabilitaire: dynamique d'un changement prosodique", in *L'accent en français contemporain*, Fónagy & Léon (eds), *Studia Phonetica* 15, 123-233, 1980.

[14] Gussenhoven, C.: On the grammar and semantcis of sentence accents. Dordrecht: Foris, 1984.

[15] Lacheret-Dujour, A., Obin, N., Avanzi, M.: "Design and evaluation of shared prosodic annotation for French spontaneous speech: from expert's lnowledge to non-experts annotations", in Proceedings of the 4th Linguistic Annotation Workshop, Uppsala, Sweden, 2010.

[16] Mesbahi, L., Jouvet, D., Bonneau, A., Fohr, D., Illina, I., Laprie, Y.: "Reliability of non-native speech automatic segmentation for prosodic feedback", Proceedings of SLaTE, 2011.

[17] Pierrehumbert, J.: *The phonology and phonetics of English intonation*, PhD thesis, published 1988 by IULC, 1980.

[18] Post, B.: *Tonal and phrasal structures in French intonation*, The Hague: Holland Academic Graphics, 2000.

[19] Speech Processing, Transmission and Quality Aspects (STQ): "Distributed speech recognition; extended advanced front-end feature extraction algorithm; compression Algorithms, ETSI ES pp. 202 212, 2005.

[20] Vaissière, J.: "Cross-linguistic prosodic transcription: French vs. English. In *Problems and methods of experimental phonetics. In honour of the 70th anniversary of Pr. L.V. Bondarko*, Volskaya, Svetozarova & Skrelin (eds). pp. 147-164, 2002.