Structuring F0 prominences (melisms): an analysis of the statistical concordances between four French speakers

Geneviève Caelen-Haumont

International Center Research MICA, CNRS UMI 2954, INPG
C10, Hanoi University Technology, 1 Dai Co Viet Str., Hai Ba Trung,
Hanoi, Vietnam.
genevieve.caelen@mica.edu.vn

Abstract

F0 prominences having the reputation deserved to vary in great proportion, most of the studies until now avoided this problem, restricting their analysis to the use of extrema (minimum / maximum), and/or the F0 range. This work presented here tried to exceed these limits: this paper aims indeed at presenting the statistical results describing the F0 prominences structuration over a corpus of 4 French female speakers.

More precisely the first part proposes a new perspective on the inner struture of the F0 prominences, and the second part focusses on the statistical results which are in concordance within the 4 speakers.

1. Introduction

Since about one century, much effort was devoted in prosody analysis and understanding, and these last years, the melodic prominences gained much more attention in scientific descriptions at the international level, particularly in speech analysis or synthesis.

The relations between form and function in the prosodic domain is a tricky problem. A lot of models have been built in order to integrate the different intonation components, for example among mostly the last decade, the IPO model [1], the British model [2], the Tilt model [3], or the PENTA model [4]. However this problem due to its great complexity remains until now unsolved.

One dimension of this complexity is that prosody fulfills several functions conveying layers of information at the linguistic, pragmatic, and subjective domains, and that each layer produces at the intra- and extra-speakers level, a wide range of forms and contours.

Arguing that these specific contours are nevertheless relevant, the aim of this present study is thus to try to clear up this profusion of forms, classifying the forms in the phonological, prosodic and tonal perspectives.

After a fast recall of the MELISM tool, the automatic segmentation and annotation it provides in the prosody domain, this paper focusses (1) on the composition of the melism structure at the phonological level, and (2) on the convergent statistical results (correlations) on the melisms components structuration, showing some laws of internal regulation.

2. About the MELISM tool

The word "melism" borrowed from the domain of singing and refering to a melodic figure spreading over the duration of the word, with a series of different notes, sometimes more important than the number of syllables in the word, is applied to speech [5], where it is related 1° to the acoustical and melodic *form* in a scalar perspective (and not binary)*,* and 2° to the subjective melodic expression [6]. A melism is caracterised by an upper melodic register and most of the time, by a great F0 contrast between two successive values.

The procedure of the automatic MELISM tool, previously named INTSMEL, was completely described in [6]. In a few words, MELISM supplies an automatic Praat TextGrid labelling. In the overall procedure, MELISM is applied to the output of the MOMEL algorithm [7] which computes targets and models F0 contour. Based on perceptive grounds, the resulting curve of MOMEL is thus similar to that found on a sequence of entirely sonorant segments.

The MOMEL procedure allows to automatically code the relevant variations of F0, under the form of successive targets which are the turning points of the modelled F0 slopes. In this perspective, the F0 curve is punctuated with labelled tones, regardless of the linguistic expression. Then the MELISM algorithm automatically codes the sequence of MOMEL target points, and this coding constitutes a surface phonological representation of tonal sequences.

Our particular purpose is to use the procedure MELISM in relation with the lexical items, which can be analysed as either isolated or integrated into the phrasal structure. Under Textgrid Praat, the context of melisms is thus manually segmented in words.

Target points automatically calculated by the MELISM procedure in the frame of each segmented word, are then grouped into a more or less complex sequence of phonological labels (for instance figure 2 below, "bb/bb/bb/bS/SH/Hc/cm/" ...) where the capitals correspond to the most acute levels targets (A, S, or H), key points of the melism definition. These phonological labels are denoting melism tones. These melism tones which occur in the limits of a word, by analogy to lexical syllables, are called *tonal syllables*.

More particularly, a set of 9 symbols (or 9 levels): acute (= 9), supra (= 8), high (= 7), elevated (= 6), middle (= 5), centred (= 4), bottom (= 3), infra (= 2), grave (= 1) is used to automatically code absolute levels corresponding to fractions (on a logarithmic scale) of the speaker's pitch range.

The present study is just concerned with the targets points of the melism words, and the relations which occur between them in the frame of each word.

3. Data-base

For the needs of the analysis, a database has been built in the frame of the PFC working group (Phonologie du Français Contemporain, Projet PFC, http://www.projet-pfc.net/),

whose objectives are to study French samples in space (all around France and other countries in the world) and time (several generations), according to the same protocol. In such conditions, we gathered recordings from 4 generations of viticultors in the same family, living since 4 generations in the same village and the same house in the South of France, near Bordeaux in the famous country of wines (Médoc). From several hours of recordings, contexts where the melisms occurred have been segmented, then the MELISM procedure applied and finally 100 melisms x 4 female speakers (thus 400 melisms in the whole) were extracted. In a second and very long task, all the data from the speakers have been very carefully checked, for instance speakers F0 extrema (F0 minima / maxima) through their different wave files and Textgrids, or the thresholds of the procedures, etc., and corrected when needed.

All the information about the 4 speakers melisms were then gathered under Excel, including the context and pauses, time (beginning, end, duration), the kind of melisms: M, non final melism (i.e. occurring inside a prosodic group), MP, final melism ending a group before a pause, MF, the same but without pause, MC, melism by contact (before/after M, MP or MF), the kind of slope of each melism tone (ascending, descending, plateau), and the semi-tone value corresponding to each target composing the melism tones.

This present study considers the overall melisms (400) of our speech data spreading from the great-grandmother to the great granddaughter. The true plateaux being very rare for these speakers, they have been reported on the rising and falling melisms, in function of the details of their internal structure. Thus it appeared that the population of the rising melisms was the biggest one, with 330 over 400 (82.5%), while the falling melisms one is only 70.
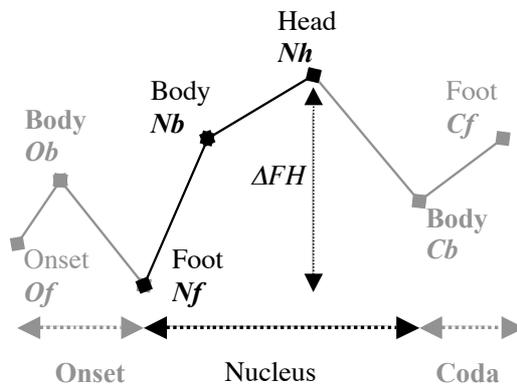

Figure 1: *The rising melism structure prototype.*

As both rising and falling melisms do occur in speech, even though their population is not the same, it means that melisms are not directly concerned with the direction of the slope: in my opinion, the slope direction is a matter of intonative space, not of melism one, running as two embedded layers.

4. The composition of the melism structure

Here is (see Figure 1 above, and Figure 2 below) the rising melism structure prototype, drawn in all the different parts it may present. The Figure 2 below shows an actual example of the prototype.

Such as a phonetic unit, this structure can be divided in 3 main parts (onset or *O*, nucleus or *N*, coda or *C)*, only the nucleus one, including the F0 minimum and maximum, being compulsory: it is indeed the heart of the melism in terms of range and direction of the slope. Nevertheless, in many cases, the onset and the coda are present.

More precisely, each of these 3 parts can be made up of several subparts, and each supbart, of one or several targets. Owing to their function, we choose the same words for the subparts: 1- *foot*, in the sense of any context anchorage, is used for the melism edges and for the lower nucleus one (the highest one being the *head*), these four targets being the most important ones in the frame of the melism, and 2- *body* for the other ones that may occur between these main targets.
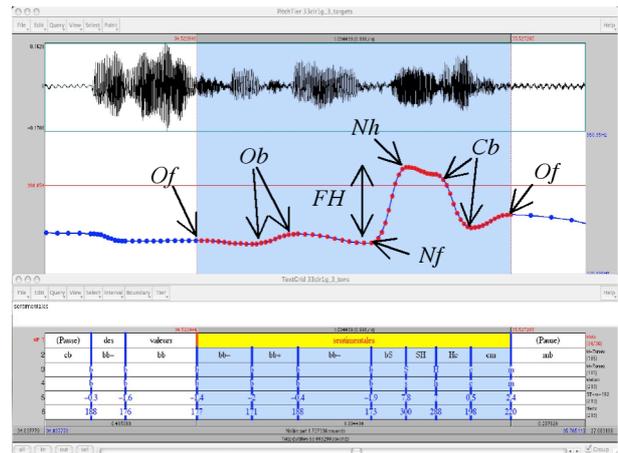

Figure 2: *An example of a MP melism (French /sentimentales/, English /sentimental/) under Praat Procedure [1]. From bottom to top: F0 values in Hz, then in semi-tones, the MELISM tonal codings, the tonal "syllables" with, in capitals, the detection of the most acute values in the melism, the manual words segmentation, and finally the F0 curve and the signal. The subparts (Of, Ob …) were manually added afterwards to the image.*

The Table 1 below presents the properties of the melism components, the number and percentages of these items. In this kind of structure, there is a clear opposition between the "body subparts" and the other subparts: as the 3 feet (onset, nucleus, coda) and head (nucleus) targets are only grounded on one value, the body ones may correspond to several ones. Moreover the body onset and body coda are often absent in our data (respectively 10% and 18%). When they are present, they assume for a better expressiveness a contrastive function vis-à-vis respectively the foot and the head nucleus.

Concerning the task of decomposing the melism structure in chunks, the first step is to detect the nucleus edges, supporting the melism range, i.e. the lowest *(Nf)* and the highest *(Nh)* F0 targets. Once done, it can be observed that between them, there can stand one or several other targets *(Nb),* the level of which being compulsorily equal to or higher than *Nf* and equal to or lower than *Nh*. As seen on the Table 1, 72% of the melisms present this middle part in the nucleus. It seems working as an aesthetic, substantial and contrastive stuff highlighting the nucleus edges using at least melody and time.

Now the next steps for decomposing the structure is the following: if a previous target *Of* stands before *Nf*, it cannot be lower than *Nf* or higher than *Nh*. If one or more targets *Ob* is/are produced between *Of* and *Nf*, in the same way than previously, its/their value can be anything except be lower than *Nf* or higher than *Nh*.
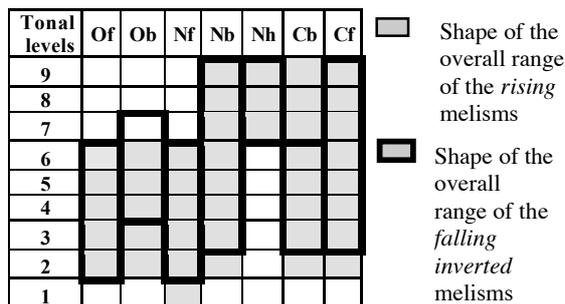
Table 1: *Parts and subparts of the melisms, properties, number and percentages (over 400 melisms).*

Melism parts	Nb	Melism subparts
195 Onset (O)	195 49%	*Of:* onset foot Only one target Optional
	38 10%	*Ob:* onset body One or more targets. Optional and often absent.
400 Nucleus (N)	400 100%	*Nf:* nucleus foot One target Compulsory
	286 72%	*Nb:* nucleus body One or more targets Optional
	400 100%	*Nh:* nucleus head One target Compulsory
271 Coda (C)	71 18%	*Cb:* coda body One or more targets Optional and often absent
	271 68%	*Cf:* coda foot One target Optional

If a last target *Cf* occurs after *Nh*, it cannot be higher than *Nh,* or lower than *Nf*. If one or more targets *Cb* is/are produced between *Nh* and *Cf*, in the same way than previously, its/their value can be anything except be higher than *Nh* or lower than *Nf*.

Comparing (Table 1 above) the percentages of the feet (onset and coda, respectively 49% and 68%), and then of the bodies (onset and coda, respectively 10% and 18%), one can remark that the second part of the melism structure gets more targets than the first one. A possible explanation could be that during the planning and production phase, a better attention paid to this part of the melism could require less targets in order to reduce the task complexity.

One can remark also that the onset and the coda parts are symmetrical, each composed of a foot (composed of one target) which is the link between the melism and its lexical /

Tonal levels	Of	Ob	Nf	Nb	Nh	Cb	Cf
9							
8							
7							
6							
5							
4							
3							
2							
1							

▢ Shape of the overall range of the *rising* melisms

▢ Shape of the overall range of the *falling inverted* melisms

intonative context, and a body (composed of one or more targets), the inner part of the onset or the coda.

Figure 3: *Targets comparison between the overall tonal patterns of 392 rising and falling melisms.*

This fact induces also another kind of symmetry, which occurs between rising vs. falling melism. The structure is just the same but put in inverted order. In order to make the rightest comparisons between them, i.e. the hardest conditions for matching them, a few targets (from 1 to 3.5% of them, concerning 8 melisms in the whole) were put away. In those conditions, the Figure 3 below shows the comparative range of the overall rising melisms (grey cases), and of the falling melisms (black frames) put in inverted order for giving them an ascending slope. Except for *Ob* and the F0 level 7 (i.e. *H* target), in the other cases, the range of the falling inverted melisms is totally embedded in that of the rising melisms.

From that point, we can draw the conclusion that merging rising and falling melisms data is allowed. In those conditions, the statistics were applied on several kind of data: rising melisms, falling melisms, then both merged (with falling structure put in inverted order to be consistent, that is to say, 400 melisms), for every sort of melisms: all the melisms (400), the melisms occurring inside a group (M), at the end of the group before a pause (MP), and then the same without pause (MF). The number of the contextual melisms (MC) remaining too small even through 4 speakers, the MC correlations are not taking into account.

5. Statistical results and discussion

For all the conditions, the analysis will consider in any case the same correlations: between respectively the Nucleus foot and Nucleus head (*Nf~Nh*), the Onset foot and the Coda foot (*Of~Cf),* and each one with respectively successively the Nucleus foot (*Of/Cf~Nf)*, the Nucleus head (*Of/Cf~Nh),* and finally with the range between Nucleus foot and head FH (*Of/Cf~FH)*.

Let us precise that we don't expect any interesting correlation between *Of~Cf*, as they both establish the link with intonational left and right contexts, neither between *Nf~Nh*, but for other arguments: if *Nf* and *Nh* would be correlated, it would mean that this melodic range would be much more repetitive and less variable than obviously it is on the melodic and time domain, and moreover, that the speaker expressiveness would be drastically reduced. As we can see on the Table 2 below, for *Of~Cf,* and *Nf~Nh*, there are, as expected, no correlations observed above 0.50.

We have now to explain the respective population of the melisms in theTable 2 below. On the left side of the table, the number indicates the overall number of the kind of melisms (for instance 146 Merged M), and on the right side, the number in brackets (for instance line 14: 0,76 (60)), indicates that 60 over 146 melisms display an *Of.*

The first observation concerns the scores quantity: the first part of the melism *(Of* target) always seems to present better results than the second one *(Cf* target).

Secondly, concerning the 70 falling melisms and the correlations between *Of/Cf~Nh*, due to their different slope direction, they behave symmetrically: the correlations above 0.50 in all the cases, occur with *Nh* and the nearest foot (as the case may be, *Of* for falling melisms, or *Cf* for rising ones). Symmetrically the inverted condition is also true: all the different kinds of melisms, except the falling ones, don't observe any *Of~Nh* correlations, neither the falling ones for *Cf~Nh*. Anyway, *Cf~Nh* correlations for all the melisms except the falling ones, and *Of~Nh* ones for the falling items, show weak scores (between 0.50 and 0.60).

That means that there exists a slight tendency that the more *Nh* is, the more *Cf* is, and conversely. On the other side, this symmetrical behaviour justifies the legitimacy of our method consisting in merging the falling and rising melisms data.

Table 2: *Correlations between the key targets of all the kinds of contextual melism, in grey colour: below 0.50, in black: above 0.50, in bold: above 0.60, and in italics: the correlations with less than 10 items.*

Correlations over :	Nf and Nh	Of and Cf
All 400 merged melisms	0,18 (400)	0,15 (130)
330 Rising Melisms	0,18 (330)	0,10 (120)
70 falling Melisms	0,29 (70)	0,40 (12)
169 Merged MP	0,45 (169)	-0,05 (53)
146 Merged M	0,38 (146)	0,39 (48)
56 Merged MF	0,27 (56)	0,36 (28)
	Of and Nf	**Cf and Nf**
All 400 merged melisms	**0,73 (195)**	0,10 (271)
330 Rising Melisms	**0,73 (173)**	0,06 (238)
70 falling Melisms	**0,66 (33)**	**0,81 (22)**
169 Merged MP	**0,62 (92)**	-0,03 (105)
146 Merged M	**0,76 (60)**	0,32 (116)
56 Merged MF	**0,81 (34)**	0,06 (45)
	0f and Nh	**Cf and Nh**
All 400 merged melisms	0,24 (195)	0,54 (271)
330 Rising Melisms	0,22 (173)	0,52 (238)
70 falling Melisms	**0,64 (33)**	0,43 (22)
169 Merged MP	0,21 (92)	0,58 (105)
146 Merged M	0,41 (60)	0,49 (102)
56 Merged MF	0,35 (34)	0,46 (45)
	Of and FH	**Cf and FH**
All 400 merged melisms	-0,50 (195)	0,24 (271)
330 Rising Melisms	-0,51 (173)	0,26 (238)
70 falling Melisms	-0,41 (33)	-0,45 (32)
169 Merged MP	-0,37 (92)	0,42 (105)
146 Merged M	-0,54 (60)	-0,05 (102)
56 Merged MF	-0,57 (34)	0,01 (45)

Another type of weak scores (between 0.50 and 0.60) is concerning *Of~FH*, i.e. the onset foot with the F0 range (ΔF0). In this case, an anticorrelation is observed, showing a tendency to the fact that the highest is *Of*, and the narrowest is ΔF0.

The *Of~Nf* correlations explain these previous scores because there exists a strongest statistical link between the feet of the onset and of the nucleus: as *Nh* is more stable in the acute levels (see Figure 2 above) than *Nf* in the low ones, we can state that the highest is *Of*, the highest is also *Nf*, which results finally in decreasing the range. This behaviour seem to be resistant to the kind of melisms, except MP which seem to be more speaker dependent: in this case, one of the four speaker shows a strong absence of correlations, making the glabal scores falling down.

These correlations are the strongest of our data, eventhough *Of* and *Nf* may be separated by one or several *Ob* targets. If we observe the *Of~Nf* correlations in the whole over the 4 speakers data, we remark that the scores often reach more than 0.80, and sometimes more than 0.90. So we can conclude that as the correlation *Of~Nf* is well established for all the melisms, the rates are depending from the speakers, where the speech ratio could be an important parameter (work in progress).

Anyway, these results mean that *Of* plays the role of an important step in the F0 prominence construction of the melism. This induces first that the melisms, if they are the expression of subjective freedom, are nevertheless linked to the intonational context (by *Of* and *Nf)*: as mentionned above, they are the junction points between the embedded layers, the intonational and melism ones. Moreover *Of* seems to supply a tonal springboard for tuning *Nf* and ΔF0*: is *Of* just a convenient foundation for the next targets in the planification task, or could *Of, Nf* and ΔF0 be also really tuned ?

The symmetrical correlations *Cf~Nh* are also existing, but as written above, the rates are weaker. This fact reinforces as well the idea that *Of* plays an important role for setting the melism nucleus.

6. Conclusions

The Praat script MELISM is a convenient tool for studying the melism structure. Through a great variability (tonality, duration, pattern …) of the melisms, this work showed that it is possible to structure a melism such as a phonetic item (onset, nucleus, coda), and more precisely into 7 subparts, where only the nucleus extrema are mandatory. Nevertheless the main targets of the melisms are also including the left and right edges *(Of, Cf)*.

These data show over the 4 speakers some invariant phenomena correlating the onset foot and the nucleus foot *Of~Nf*, eventhough the rate (ever above the treshold of 0.50), may be more dependent from the speaker. The correlations seem to be stronger when occurring between the targets of the initial part of the melism, suggesting that the speakers pay more attention to their production than later. In this perspective, the first melism target, *Of,* seems to play an important role in the melodic setting up and expansion of the melism.

The next work bears upon the speaker dependent data, and upon the time duration.

7. References

[1] J. 't Hart, R. Collier and A. Cohen. 1990. *A perceptual Study of Intonation — An experimental phonetic approach to speech melody.* Cambridge University Press, Cambridge, England.

[2] A. Cruttenden. 1997. *Intonation*. Cambridge University Press, Cambridge, England.

[3] P. Taylor. 2000. Analysis and synthesis of intonation using the Tilt model. *Journal of the Acoustical Society of America,* 107, pages 1697-1714.

[4] Y. Xu. 2004. The Penta Model Of Speech Melody: Transmitting Multiple Communicative Functions In Parallel. Proceedings of From Sound to Sense: 50+ years of discoveries in speech communi-cation, pages 91-96, Cambridge, MA.

[5] G. Caelen-Haumont and B. Bel. 2000. Le caractère spontané dans la parole et le chant improvisés: de la structure intonative au mélisme. *Revue Parole*, pages 251-302.

[6] G. Caelen-Haumont and C. Auran. 2004. The Phonology of Melodic Prominence: the Structure of Melisms, Proceedings of Speech Prosody 2004. Nara, Japan, pages 143-146.

[7] D. Hirst and R. Espesser. 1993. Automatic labelling of fundamental frequency using a quadratic spline function. *Travaux de l'Institut de Phonétique d'Aix,* 15, pages 71-85.

[8] P. Boersma and D.J.M. Weenink. 1996. Praat, a system for doing phonetics by computer. Amsterdam: Institute of Phonetic Sciences of the University of Amsterdam.