coustics'08
Paris
June 29-July 4, 2008
www.acoustics08-paris.org

Melodic prominences structures: exploring to what extent the speaker variability is spreading

Geneviève Caelen-Haumont

MICA Center, C10, Hanoi Universirty of Technology, 1 Dai Co Viet Str., Hai Ba Trung,
Hanoi, Hanoi, Viet Nam
genevieve.caelen@mica.edu.vn

Melodic prominences in speech have a deserved reputation of conveying a great part of variability, as they are greatly based on subjective impulses or feelings. We think that these specific contours are nevertheless relevant, and that their structure, while obeying to internal laws of regulation, nevertheless leaves room for an inter-speakers variability. Generally no specific tool was used to describe these F0 contours with precision. An automatic analysis tool MELISM was then developed allowing accurate descriptions of F0 salience (melisms). This paper aims at 1° describing the melisms structures 2° presenting the main results over 4 speakers exploring to what extent the speaker variability can spread. The melism structures are explored and analysed reducing the whole variability to three main components (onset, nucleus, coda) and their respective subparts, where only the nucleus is the compulsory part. Statistics running on these structures put to light a set of interesting laws about the internal regulation of melisms but on the other side, the part devoted to the speaker variability.

1 Introduction

These last years, the melodic prominences gained much more attention in scientific descriptions at the international level, in speech analysis or synthesis. A lot of models have thus been built in order to take into account the different intonation components, for example among the most recent ones, the IPO model [1], the British model [2], the Tilt model [3],or the PENTA model [4].

Up to now however, many problems remain unsolved in the domain of prosody, such as for instance in a general perspective, the complex relations between form and function as prosody fulfills several functions at the linguistic, pragmatic, and subjective levels, using even the same acoustic features, and producing moreover a great number of forms. Another tricky problem concerns as well the domain of emotions, and especially the means to prosodically distinguish between very active / very positive emotions (in reference to the FEELTRACE orthogonal axes [5]) and the very active / and very positive personal investment in speech. This personal involvement takes place in a single or at the most two words, giving to it specific charateristics (especially high F0 values and a great range). Such subjective prominences were called *melisms* [6]. 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 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.

The aim of this study is to clear up the melism structure among the profusion of the forms. Though these structures supply a high level of variability, we think that these specific contours are nevertheless relevant, and that their structure obeys to internal laws of regulation.

Grounded on a databank of 400 melisms (100 x 4 speakers), this paper focuses on the melism melodic shape, its archetypal structure, and with the support of statistical correlations, the internal regulation of melisms but on the other hand, the part devoted to the speaker variability.

2 The MELISM tool

The procedure of the automatic MELISM tool, previously named INTSMEL, being completely described in [7], we just precise that MELISM supplies an automatic Praat TextGrid labelling. In the overall procedure, MELISM is applied to the output of the MOMEL procedure [8], which computes targets and models the F0 contour. 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.

The Figure 1 below presents from bottom to top: the 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 arrows indicating the tonal syllables and the tonal targets, and the subparts (Of, Ob …, see below the section 4) were manually added afterwards to the image.

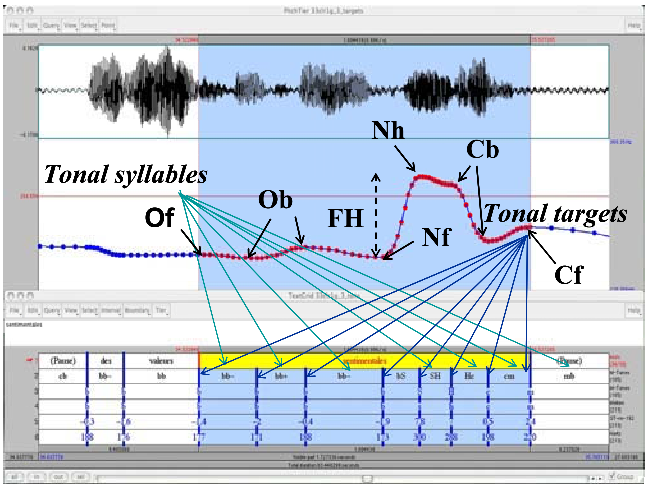

Figure 1 : An example of a MP melism (French /sentimentales/, English /sentimental/) under Praat procedure [9.]

2.1 Segmentation, automatic annotation and labeling

Our particular purpose is to use the MELISM procedure in relation with lexical/grammatical items. Under the Praat Textgrid, the linguistic context of melisms is thus manually segmented in words.

The 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 1 above, /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 between 2 tonal targets are denoting melism tones. These melism tones which occur within the limits of a word, by analogy to lexical syllables, are thus called *tonal syllables*.

In particular, a set of 9 symbols: 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 each speaker's pitch range.

3 Database

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 (Cussac-Fort-Médoc) and the same house in the South of France, near Bordeaux. From several hours of recordings, contexts where the melisms occurred have been segmented, then the MELISM procedure was 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 if necessary.

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 with 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 tone.

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 329 over 400 (82.3%), while the falling melisms one is only 71.

As both rising and falling melisms do occur in speech, even if 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 Internal structure of the melisms

Such as the syllables and phonetic units, a melism is composed of an initial part (Onset or O), a central part (Nucleus or N), and a final part (Coda or C). The Figure 2 below presents the rising melism structure prototype, drawn in all the different parts it may supply, and the figure 1 above shows an actual exemple of the French melism *sentimentales (sentimental* in English).

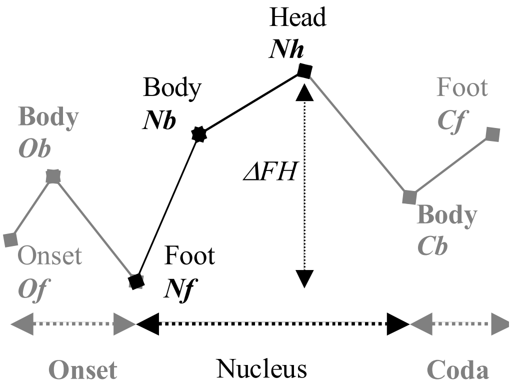

Figure 2: The rising melism structure prototype.

Among these 3 parts, *only the nucleus one, including the F0 minimum and maximum, is compulsory:* it is indeed the *heart of the melism* in terms of range and slope direction. 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. The subparts assuming the same fonction are denoted by the same terms, for instance for the rising melisms:

1- "*foot"*, in the sense of any *context anchorage*, is used for the beginning *(Of)* and the end *(Cf)* of the linguistic (i.e. lexical and more rarely grammatical) unit, and for the beginning of the nucleus part *(Nf)* which can be also the melism beginning. With the nucleus head *(Nh)* which is the melism highest F0 value, *these four targets (Of, Nf, Nh, Cf) are the most important ones in the frame of the melism,*

2- "*body"* is used for the other ones that may occur between these main targets. The different bodies along the melisms parts seem working as an aesthetic, substantial and constrastive stuff highlighting the nucleus edges, using at least melody and time.

The Table 1 below presents the properties of the melism components, the number and percentages of those items. The population and percentages are computed on all the rising and falling melisms, as they occur in the speech. In this kind of structure, there is a clear opposition between the "body subparts" and the other subparts: the 3 feet (onset, nucleus, coda) and head (nucleus) targets are indeed only grounded on one value, whereas the body ones may correspond to several ones. Moreover the body onset and body coda are often absent in our data (respectively 12% and 15% over the 400 melisms). When they are present, they assume for a better expressiveness a contrastive function vis-à-vis respectively the foot and the head nucleus.

One can remark also that the onset and the coda parts are symmetric, each composed of a foot (composed of one target) which is the junction between the melism and its lexical / intonative context, and a body (composed of one or more targets), the inner part of the onset or the coda.

This fact induces also another kind of symmetry, which occurs between the rising vs. falling melisms. The structure is just the same but puts in the inverted order. After examinating in details the F0 targets of the rising and falling melisms, we observed that the range of the falling inverted melisms is totally embedded in that of the rising melisms [10].

Melism parts	**Population and percentages**	**Melisms subparts**
208 Onset *(O)*	**208 (52%)**	***Of:* Onset foot** **Only one target** **optional**
	46 (12%)	***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 and often present**
	400 (100%)	***Nh:* Nucleus head** **One target** **Compulsory**
259 Coda *(C)*	**61 (15%)**	***Cb:* Coda body** **One or more targets** **Optional and often absent**
	259 (65%)	***Cf:* Coda foot** **One target** **optional**

Table 1: Parts and subparts of the melisms, properties, number and percentages (over 400 rising and inverted falling melisms).

From that point, we can draw the conclusion that merging rising and falling melisms data is not a problem. In these conditions, the statistics were applied on several kinds of data: rising melisms (R), falling melisms (F), then both merged (i.e. 400 RF melisms, with falling structures put in the inverted order to be consistent, such as the rising structures), and 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 just reported but not took into account.

5 Speakers statistical correlations

First of all we have to explain the respective population of the melisms in the table 2 below. On its left side, the number indicates the whole population of the kind of melisms (for instance 329 rising melisms or R), and on the right side, the upper and bottom numbers (for instance: ***0,73*** 172), indicate that 172 over 329 melisms display an *Of* achieving a *Of/Nf* correlation rate of 0.73.

Before any comment of the results, let us precise that we don't expect any interesting correlation between *Of / Cf*, as they both establish the junction with left and right intonative contexts. In the same way, we don't expect for the rising melisms any correlation 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.

This table 2 below shows obviously that for the 4 speakers:
1° the strongest correlations rates occur in the first part of the melisms,
2° that *Of/Nf are correlated.* We have to keep in mind that these two feet can be separated by one or several F0 targets,
3° the symmetric correlations in the melism space of *Of/Nf* are *Cf/Nh.* For them, there exists also correlations but their rates are lower,
4° for the 71 falling melisms (F), *Cf/Nf* presents also a strong correlation rate (0,81), as these 2 targets are symmetric to the *Of/Nf* ones for the rising melisms which present as just seen above, also good rates,

Correlations	Nf and Nh	Of and Nf	Of and FH	Of and Nh	Cf and Nf	Cf and FH	Cf and Nh	Of and Cf
400 R + RF	0,19 400	***0,74*** 195	***-0,52*** 195	*0,24* 195	0,12 271	0,22 271	***0,55*** 271	0,14 130
329 R	0,19 329	***0,73*** 172	***-0,53*** 172	0,21 172	0,08 238	0,24 238	***0,53*** 238	0,08 120
71 F	0,21 71	***0,66*** 34	-0,43 34	0,08 34	***0,81*** 23	-0,44 23	0,23 23	0,40 12
167 R + RF MP	0,14 167	***0,63*** 90	0,03 90	0,20 90	-0,03 103	0,43 103	***0,59*** 103	-0,04 52
147 R + RF M	0,39 147	***0,78*** 63	***-0,58*** 63	0,49 63	0,34 117	-0,09 117	0,49 117	0,35 51
56 R + RF MF	0,25 56	***0,78*** 34	***-0,53*** 34	0,40 34	0,25 45	-0,01 45	0,44 45	0,40 28

Table 2: Four speakers statistical correlations between the main targets of all the melism kinds in a contextual perspective, with in bold italics, the correlations over 0.50.

5° *Of/FH* presents also some anti-correlations above the -0.50 threshold, which means that there is a tendency that higher *Of* is, and narrower is the F0 range *FH.* This tendency is not surprising, because first *Nh* spreads only on 3 grades (/H, S, A/), and secondly *FH* is calculated on the range the base of which is *Nf,* correlated with *Of*,
6° an absence of correlations provides an interesting information / confirmation: while *Nf/Nh* are corresponding to the inner part of the melisms -the only compulsory ones with both the lowest and the highest F0 values of the melisms-, they are not correlated at all, nor with any previous subpart of the melism. This front open space suggests and confirms as well, that *the highest F0 value as well as the lowest one, are left to the speaker freedom expression.*

6 Melisms variability

This section is concerned with the extent of the melism variability.

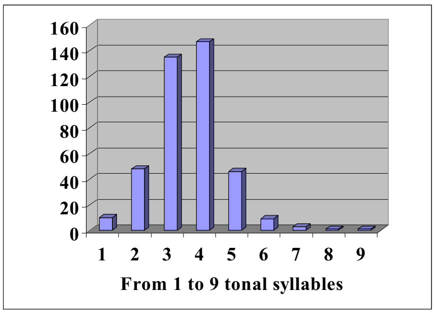

Figure 3: Number of the melisms by the number of the tonal syllables (over 400 melisms, 4 speakers).

The Figure 3 above shows the number of melisms by the number of the tonal syllables in the 400 melisms. The most numerous melisms occur with 2 to 5 tonal syllables, and mainly with 3 and 4 ones (71% of the overall melisms).

The figure 4 below presents the number of the melisms by the 9 tonal targets. For the highest range, the target *H* is the most frequent one, while for the lowest ones, the targets *b* and *c* have the best rates. These two facts put together show that the most numerous melisms are focussed on the most central targets. It is not surprising between it is the passing place of the F0 slopes, and it needs the weakest effort in comparison with the lowest and the highest ones.

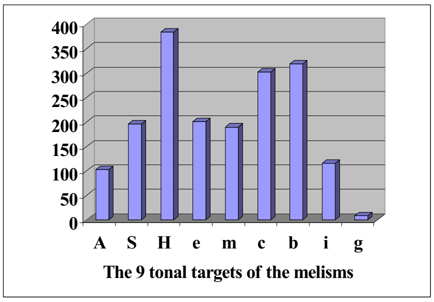

Figure 4: Number of the melisms by the 9 tonal targets (over 400 melisms, 4 speakers).

The Figure 5 below presents a distribution of the 400 melisms over the 4 melism contextual categories. The MP ones (before a pause) is the main category (42% of all of them), but this global result hides some disparity among the speakers (from 33 to 59%), the youngest speakers producing less melisms before a pause than the eldest ones. Concerning the M melisms which occur inside a group, they represent in average 37% of the total.

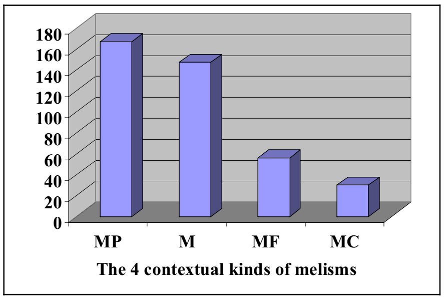

Figure 5: Number of the melisms in each kind of melism.

These two categories, MP and M, being the most important, we analyse now in more details each speaker contribution. We observe that the four speakers achieve a sort of balance in the melisms distribution. The eldest speaker (HV1) and the youngest one (LR2) show an equal distribution (respectively 40-40% and 33-33%), while the two other ones, in the middle age, balance their rates between these two categories: the elder one (SP1), reduces their number from MP melisms to M ones (MP: 59% to M: 23%), whereas the younger one (LR1) increases them (MP: 35% to M: 51%).

Moreover the balance is completed with the MF melisms, denoting the melisms ending a group without a pause, which represent only 14% of all the melisms. As a matter of fact, for the 3 eldest speakers (HV1, SP1, LR1), the rates for the MP, M, and MF melisms are exactly equal (94%), the MC melisms being necessarily also equal (6%). On the other hand, the youngest one (LR2) presents the smallest dispersion of her items (MP: 33%, M: 33%, MF: 22%, MC: 12%).

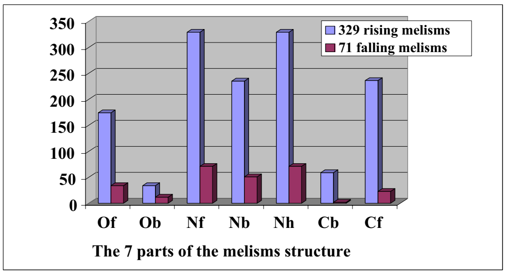

Figure 6: Number of the melisms in each of their 7 subparts.

The Figure 6 above shows a distribution of the 400 rising and falling melisms over each subpart of the melism structure. We can observe that 1° globally the distribution of the two kinds of melisms according to the ascending vs. descending direction of their slopes, follow the same pattern, 2° the number of items is quite different from one subpart to the another one 3° the most numerous subparts are the central ones, that are the only compulsory ones.

7 Inter-speaker variability

We consider now the inter-speaker variability through the best and more consistent correlations, i.e. which exceed 0.50. Concerning the best ones *Of/Nf* reaching on the whole 0.74 (4 speakers, 195 melisms over 400 presenting an *Of*), the figure 7 below shows to what extent they are stable among the speakers. For a better visibility, we take only into account the merged rising and inverted falling melisms.

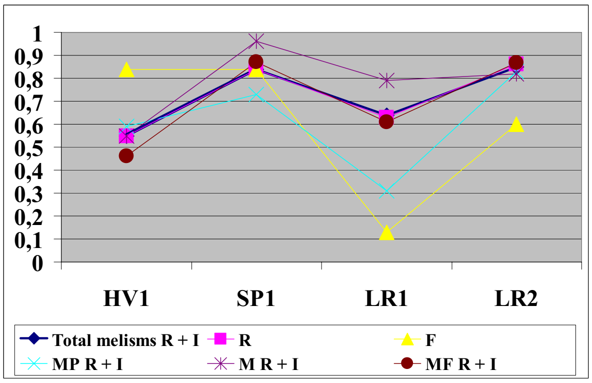

Figure 7: Inter-speakers correlations between *Of* and *Nf* melisms targets.

First of all, we oberve that the falling melisms (Δ symbol) behavior in a very different way than the others kinds of melisms, except for SP1. Of course in that case, we took into account the actual falling structures, not the inverted ones. Anyway on the whole we note that the global 4 speakers behaviors concerning the rate increases / decreases tend to covary. Apart from the falling melisms, which in fact can better be checked with the symmetric targets correlations (*Nh/Cf,* see below the following figure), we remark that 1° concerning the narrowest rates brackets, HV1 and LR2 are the most concerned with it, SP1 a little bit less, whereas LR1 supplies a great variability, especially for the MP rising and inverted falling melisms, i.e. the melisms ending a group, 2° concerning the rates values, SP1 and LR2 present the best ones, and especially SP1 with a 0.96 rate for the M melisms (but only 10 items). Though the HV1 rates are well grouped, they don't exceed 0.59, which is very low. As for LR1, her rates bracket is very large, from 0.28 to 0.80, 3° *in the field of the highest consistent rates, the M category of melisms, inside the groups, is first*. We have to note that M is *the best expression of affectivity,* as it cannot be questionned on the matter of exercising a syntactic function, which is not the case for the other kinds of melisms (MP and MF) where the syntax constraints can possibly lead to less efficient rates.

The second correlations in discussion are the *Nh/Cf* ones, the symmetric targets in the melism structure. In the figure 8 below, we remark first that the variability is much greater, and that the rates are globally lower.

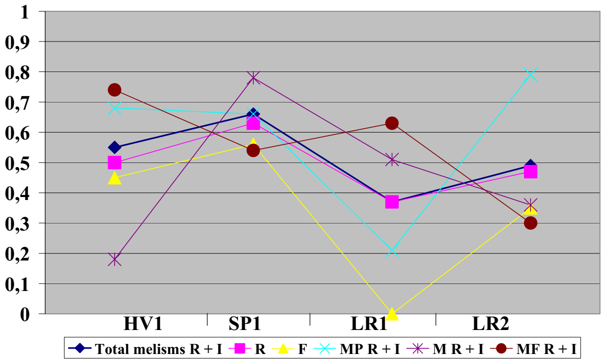

Figure 8: Inter-speakers correlations between *Nh* and *Cf* melisms targets.

Then we remark that 1° concerning the narrowest rates brackets, the same speakers as before as concerned with it, namely HV1, SP1 and LR2, but with some exceptions (M melisms), 2° concerning the rates values, HV1 and SP1 present the best ones. It is worth noting for the fore correlations melisms (*Of/Nf* examinated above) that the HV1 rates were low, whereas for these ones, sometimes they are surprisingly better (from 0.5 to 0.74), especially the MP and MF ones: indeed whereas the MF *(Of/Nf)* were not reaching 0.50*,* now it reaches 0.74. In those conditions we can state that HV1 presents another sort of speech strategy for the melisms production consisting in correlating the final melisms targets rather than the fore ones. For the MP correlations, we remark the same process, and the rates increase from 0.59 (*Of/Nf*) to 0.68 (*Nh/Cf*). Of course these two rates are not so good, but the speaker HV1 settles both correlations at once in the melisms frames. Besides, the M correlations considerably vary among the speakers, from 0.18 to 0.78. SP1 supplies a high rate (0.78) though her *Of/Nf* rate was yet excellent (0.96), so she succeeds in correlating her fore targets between them as well as the final ones, which is very rare because of the task effort. We remark also the same double correlations for the MP melisms with SP1 and LR2. 3° In the whole for the *Nh/Cf* rates, we don't note any consistent rates for the highest rates through the speakers, 4° Obviously LR2 uses the strategy of correlating only the melisms fore targets.

Considering that the speech flow could be an important parameter enabling to explain these results, indeed SP1 supplies the lowest speech flow with pauses (4.04), while LR2 presents the lowest average number of syllables by the number of pauses [10].

8 Conclusions

This paper aims first at clearing out the prominences / melisms melodic frame by 1° defining a prototype structure 2° grounding it on a statistical correlations evidence. For a much larger study, please see [10]. Secondly, it supplies a study on the domain of the variability, as well in the field of the melisms structure than that of the inter-speaker one. It was shown in particular that the speakers can use several strategies to structure the melodic melism space, which can explain a consistent part of the inter-speakers variability.

9 References

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

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

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

[4] Y. Xu, 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 (2004).

[5] R. Cowie, E. Douglas-Cowie, S. Savvidou, E. Mcmahon, M. Sawey, M. Schröder, FEELTRACE: an instrument for recording perceived emotion in real time, *Proceedings of the ISCA Workshop on Speech and Emotion*, Belfast, http://www.qbc.ac.uk/en/isca /proceedings (2000).

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

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

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

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

[10] G. Caelen-Haumont, Emotion, emotions and prosodic structure: an analysis of the melisms patterns and statistical results in the spontaneous discourse of 4 female speakers over four generations, Peter Lang (to be published).