HAL open science

The phonology of melodic prominence: the structure of melisms

Geneviève Caelen-Haumont, Cyril Auran

To cite this version:

Geneviève Caelen-Haumont, Cyril Auran. The phonology of melodic prominence: the structure of melisms. Speech Prosody, Mar 2004, Nara, Japan. pp.143-146. hal-00256390

HAL Id: hal-00256390

https://hal.science/hal-00256390v1

Submitted on 15 Feb 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The Phonology of Melodic Prominence: the Structure of Melisms

Geneviève Caelen-Haumont[1], *Cyril Auran*[1,2]

1- Laboratoire Parole et Langage / Université de Provence, Aix-en-Provence, France
2 - IUFM d'Aix-Marseille, Aix-en-Provence, France
gcaelen@lpl.univ-aix.fr,cauran@wanadoo.fr

Abstract

This paper aims at proposing a surface phonological tonal annotation and stylization fitted to the lexical space. More precisely it makes it possible to phonologically structure the F0 variations in prominent words. In previous studies, this specific F0 configuration in such words has been called *melism*. These principles are integrated in an automatic procedure (INTSMEL) which supplies an automatic Praat TextGrid labelling. In the overall procedure, INTSMEL (and/or INTSINT) is applied to the output of the MOMEL algorithm which computes targets and modelled F0 contour. INTSINT and INTSMEL have complementary goals: the former is devoted to the annotation of intonation, the latter to the (prominent) word (or suite of words) annotation. The aim of this paper is to describe this annotation method, previously to its exploitation and evaluation in forthcoming papers.

1. Introduction

The question of melodic or prosodic annotation and stylization is very important and since the origin of intonation studies, many efforts at the international level have been devoted to such a task. In fact, studies in this domain can be divided into two main approaches whether they integrate or not in their formalism a relation to the linguistic frame, and especially a reference to syntactic phrase boundaries. Such a perspective for a single language is no doubt richer and more descriptive, but the major problem is its adequacy to reality: Generality often mismatched with local reality. On the other hand, a perspective which aims only at describing F0 variations make fewer mistakes and can moreover be applied to a large panel of F0 systems in the world.

2. F0 range, prominence, pattern, focus and melism

Numerous studies have been devoted to characterizing acoustic modifications of F0, related to the expression of a linguistic meaning in affective conditions of speaking. More precisely, concerning the matter of F0 amplitude, various denominations can be encountered in papers. In fact none of them seems to be convenient. For instance, the term *prominence* (o*r salience)* is very imprecise and refers to an impressive meaning. The term *range* refers to F0 amplitude but delivers no information about the range's shape. The term *pattern* is generic, and does not refer to a phonological perspective. As to the term *focus*, it is probably one the most common words used in this domain, but its meaning is still inadequate: 'focus' refers to a binary process (focussed / unfocussed) whereas it actually is, as has been often noted, scalar. Moreover, it is rather fuzzy as it induces confusion between the acoustic, semantic and pragmatic domains.

In these conditions, it was necessary to find a word 1° related to the acoustical and melodic *form* 2° expressing the notion of a structured shape with an adapted granularity 3° in terms of phonological structure, i.e. of a tonal system. To convey such a definition, the term of *melism* appears conducive [1]. It is borrowed from the domain of singing and refers to a melodic figure spreading over the duration of the word, with a suite of different notes, sometimes more important than the number of syllables in the word.

3. MELISM: an automatic system of tone annotation

3.1. MOMEL, INTSINT and INTSMEL

The MELISM procedure actually consists of three chained "main" algorithms (MOMEL, QSP and INTSMEL) and two "minor" tools for Praat TextGrid and PitchTier format conversions.

The MOMEL algorithm, to begin with, aims at modelling the actual F0 curve so that any microsegmental characteristics (the *micro-prosodic component*) should be factored out [3]. The resulting curve is thus similar to that found on a sequence of entirely sonorant segments and constitutes the *macro-prosodic component* ([6]), [7]).

The processing of the quadratic spline functions is accomplished through the QSP algorithm: sequences of MOMEL target points (within a time / frequency space) are taken as input values, and F0 spline-modelled values are computed every 10 ms. for the entire speech signal.

The combination of MOMEL and QSP therefore allows us to treat a sequence of target points as an appropriate phonetic representation of F0 curves.

The INTSMEL algorithm, though generating an output visually quite similar to that produced by INTSINT, actually diverges with it both in its theoretical bases and goals.

On the one hand, the INTSINT algorithm automatically codes the sequence of MOMEL target points using a limited set of abstract tonal symbols {M, T, B, H, L, S, U, D} standing for *Mid*, *Top*, *Bottom* (absolute tones), *Higher*, *Lower*, *Same*, *Upstepped* and *Downstepped* (relative tones) respectively ([5]). The INTSINT coding constitutes a surface phonological representation of intonation independent from any a priori phonological inventory of the intonation patterns of a given language.

On the other hand, the INTSMEL algorithm codes the sequence of MOMEL target points taking into account the principles detailed in the following sections. More particularly, a set of 9 symbols ({a, s, h, e, m, c, b, i, g}) is used to code absolute levels corresponding to fractions (on a logarithmic scale) of the speaker's pitch range. Target points are then

automatically grouped into melism tones (For INTSINT, MELISM and INTSMEL, see section 6 below).

3.2. Praat and MELISM

All the algorithms related to the implementation of MELISM are called in a modular way within a single Praat script called melism.praat; the procedure, which offers batch processing functionalities, can be divided into four main stages:

First, F0 extraction is processed using an accurate autocorrelation method with pitch floor and pitch ceiling values either given by the user or set to classical 75Hz-600Hz values. F0 values (within a time/frequency space) are then saved to ASCII files for MOMEL processing.

Secondly, the MOMEL binary computes target point values from the extracted F0 values. The output takes the form of an ASCII file. A Perl script then converts this output into PitchTier format for further use within Praat.

Thirdly, taking the MOMEL target files as input, the QSP binary generates other ASCII files with F0 values corresponding to the spline-modelled F0 curve. This output is then converted into Praat PitchTier format for subsequent visualization and editing by the user.

Finally, the actual coding of the target points and their grouping into melism tones are computed using dedicated procedures within the melism praat script. A TextGrid file is eventually generated for subsequent use within Praat.

4. The objective of melism annotation

4.1. The acoustic properties of the melism

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. The MELISM procedure then computes the resulting tones.

This procedure is convenient to some specific research. Our particular purpose is to use it in relation with lexical items, which can be analysed as either isolated or integrated into the phrasal structure. After the segmentation phase, each linguistic item is coded from the left boundary to the right one with a sequence of phonological labels. This sequence can be either simple (only one slope, ascending or descending) or complex (an alternation of opposite or parallel slopes or plateaux), but, by definition, in every case the structure of the melism begins or finishes in a very high register.

More precisely the prosodic correlates of a given melism are 1° a large F0 excursion (internal, or external if an F0 break occurs within the previous / following word), 2° the implication of at least the infra-acute level (symbolised by level *s* in our tonal system) or more (level *a),* or conversely, level *h*, if it involves a significant F0 excursion (about 10 semi-tones), all modulations which generally, but not systematically, accompany a clear decrease in speaking rate, and eventually, a dramatic increase in energy.

4.2. The grounds for this annotation

As previously mentioned, this system of annotation is grounded, as the INTSINT annotation, on the MOMEL automatic procedure [5]. The MOMEL procedure allows to rebuild (and moreover to compute the F0 value of) the speaker's underlying pitch targets which are not always reached, and which defy classical F0 processing methods since they may occur on unvoiced parts of the signal (unvoiced segments or pauses). In any case nevertheless, the objectives of this phonological annotation using INTSMEL are:

- to give a quantitative information (F0 value) on the speaker's targets,
- to propose on these groundings an annotation system and a stylisation independent from the phrase, sentence, corpus, speaker, gender or age parameters,
- to relate such a coding to F0 absolute levels according to the syntagmatic axis of the utterance, but relative levels according to the maximal speaker's melodic register,
- to phonologically describe the internal structure of melismed words,
- to compare their different patterns in function of their syntactic status, their semantic or pragmatic values,
- to compare the different melodic systems existing in the different languages, including tonal ones,
- to objectively test theorical hypotheses in experimental procedures.

5. The phonological system

5.1. Comparison between different annotation systems

There exist numerous prosodic annotation systems for speech analysis. The aim of this paper is not to supply a large appraisal of these different systems, but to point out the specificity of the INTSMEL coding with some well-known language dependent vs. independent systems.

One of them is the ToBI procedure [8]. It proceeds from a theoretical point of view, but also depends on the describer's empirical judgements. It appears as a mixture of perception (break indices) and acoustic and phonological perspective (pitch events such as boundary tones and pitch accents); finally, it is directly linked to the linguistic analysis of a given language, since the symbols take into account stressed syllables, phrase junctures, and initial and final sentence boundaries.

The PIT system [2], following the line of research initiated in the mid sixties at IPO Institute [4] for speech synthesis, has been developed within the scope of an automatic analysis of French intonation using an automatic speech recognizer (ASR). It is grounded on a primary segmentation of the speech signal (and then of the F0 contour) into syllable-sized segments, and iteratively on a second segmentation into successive tonal segments, including a perceptual integration of short and mid-term pitch variation, glissando threshold and differential glissando threshold. At this stage, the system supplies tonal segments, stylised contours *(tonal score),* and its outputs enters the synthetic speech module. Partially implemented in the fully automatic system, the semi-automatic Mingus system, relying on Piet Mertens' [9] tonal model of French intonation, proposes a set of 4 tones x 2 (accented syllables vs. unaccented ones), which describe absolute F0 variations greater than major third, and relative variations taking place within the interval of a major third ("downstepped" and "upstepped" tones). As the ToBI system, the PIT annotation procedure relies on prior linguistic analysis, and especially, syntactic analysis of the studied language.

Rossi and Chafcouloff's work [10], one of the major references for french studies, constitutes a source for the INTSMEL system. This system displays 6 intonation levels,

covering the whole of the speaker's range, and details the corresponding quantitative values and their standard deviations. 6 *intonemes* are then defined with respect to the speaker's mean F0, as a function of form (a specific height, a specific orientation of the slope, a specific F0 range) and syntactic content. These tonal morphemes enable the setting up of contrastive pairs (question vs. assertion; minor (or listing phrase) vs. major continuative phrase; minor vs. major conclusive phrase), playing specific syntactic functions. The INTSMEL system differs on the one hand in the number of levels, and, on the other hand, on the nature of syntactic functions, the only items observed here being lexical.

Nevertheless some other systems are independent from any prior linguistic analysis, and can consequently be applied to any language (tones language or not) and any type of melodic organisation; such analyses, indeed, were performed using the INTSINT system [5].

At the output of MOMEL procedure, both systems, INTSINT and INTSMEL provide an automatic F0 annotation. The difference between INTSINT and INTSMEL lays upon 2 dimensions: the principle of computation of the target and the reference to the linguistic items.

As to the first point, the INTSINT annotation splits up the tones between 1° the absolute ones which are Top *(T)*, Mid *(M),* Bottom *(B),* calculated on the speaker's maximum range and medium value, and the relative ones, Higher *(H),* Same *(S*), and Lower *(L),* computed in relation with the previous target, whatever it may be, absolute or relative, and 2° two other specific relative tones, Upstepped *(U)* and Downstepped *(D)*. These last tones allow to label a sequence of tones with a more reduced interval.

The INSTMEL annotation considers only an absolute coding in so far as it acts in the limited frame of a word, observed in isolation or not. Thus these two conceptions are related to each other, the former (INTSINT) providing an annotation of intonation, regardless of the linguistic nature of the annotated string, and the latter (INTSMEL), an annotation of melisms within word boundaries. Perceptual equivalence was found between on the one hand resynthesized signal using PSola procedure and MOMEL-modelled F0 values, and, on the other hand, the original signal ([6], [7]).

The INTSMEL annotation thus appears as a system devoted to a specific goal: annotating the F0 variations, and then phonologically describing the tonal structure of the prominent words. Under this consideration, INTSINT and INTSMEL can be regarded as systems providing complementary annotations.

6. INTSMEL

6.1. The speaker's range and the tone levels

Since Delattre 1966, the speaker's range is usually divided into 4 levels, but a precise study of the shapes of melisms requires greater precision for the description of F0 variations. For instance, the problem of the neutralisation of F0 variations is important. According to the INTSMEL procedure, such neutralisation is implemented within a span of a fourth of a level above and below each tonal boundary (i.e.one semi-level). The 5 boundary tones, with their respective neutralisation ranges, eventually lead (see Figure 1 below) to 9 equal levels (logarithmic scale).

A function exists however between the initial 4 and final 9 levels, a conversion can easily be made (for instance E = 1/2 h + e + 1/2 m, and e = 1/2h + 1/2m= 50% of E). In our perspective, the minimum and maximum F0 values are extracted from the whole of a given speaker's data; the accuracy of the coding thus depends on the amount of data exploited.

To give a more accurate description, we make a distinction between intra-level variations, and plateaux. By opposition to the plateaux which, by definition, are not oriented, intra-level variations are always annotated with the symbols +/- (eg.: *mm+; ss-,* etc.), which codes the existence of an ascending vs. descending slope. The targets are naturally graduated (a > i), but in a sequence of tones, the same target level may occur several times. A capital letter thus avoids any ambiguity: < *hh- he eh hH* >, < *Sm, mm, ms* > and so on. The example < *ss sA as sS* > indicates 1° that sA > as, 2° sS > ss, (and of course that 3° sA > sS).

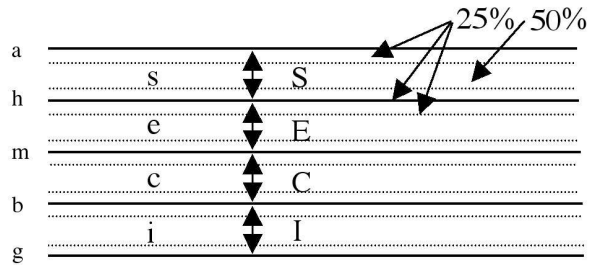

Figure 1. *The 4 and 9 F0 levels for melisms, with a: aigu/acute, s: supérieur/supra, h: haut/high, e: élevé/elevated, m: moyen/mid, c: centré/centred, b: bas/bottom, i: inférieur /infra, g: grave/grave, and I, C, E, S: respectively, 1st, 2nd, 3rd and 4th F0 level.*

6.2. Description of the tones for melisms

The combination of these 9 levels results in 81 tones. These 81 tones can render any F0 configuration in words, but for melism description, only 45 tones are considered useful.

ton	aigu	supérieur	haut	élevé	moyen	centré	bas	inférieur	grave
	a	**s**	**h**	**e**	**m**	**c**	**b**	**i**	**g**
a	***aa***	**as**	**ah**	**ae**	**am**	**ac**	**ab**	**ai**	**ag**
s	**sa**	***ss***	**sh**	**se**	**sm**	**sc**	**sb**	**si**	**sg**
h	**ha**	**hs**	***hh***	**he**	**hm**	**hc**	**hb**	**hi**	**hg**
e	**ea**	**es**	**eh**	*ee*	em	ec	eb	ei	eg
m	**ma**	**ms**	**mh**	me	*mm*	mc	mb	mi	mg
c	**ca**	**cs**	**ch**	ce	cm	*cc*	cb	ci	cg
b	**ba**	**bs**	**bh**	be	bm	bc	*bb*	bi	bg
i	**ia**	**is**	**ih**	ie	im	ic	ib	*ii*	ig
g	**ga**	**gs**	**gh**	ge	gm	gc	gb	gi	*gg*

Table 1. *Matrix of tones used for the description of F0 configurations in words, and especially for melisms. The bold and italic characters correspond respectively to melism tones and to plateaux (linked to melisms or not).*

In our own experiments, we restricted the number of distinctive tones to a combination of 3 levels, *a, s* and *h.* Table 1 below presents the 81 tones, and among them, the 45 used in the description of melisms are in bold; the italic

characters correspond either to intra-level variations or to plateaux (i.e. F0 variations which occur below the 50% threshold of the 9 levels, corresponding in fact to a neutralisation of the variation).

Figure 2 below completes the previous one and gives an illustration of the descending tones and of the plateaux related to melims, and their ranges:

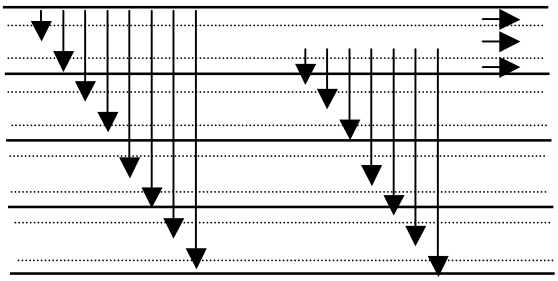

Figure 2. *Example of the range covered by the 15 descending tones in the melisms* (*with a: aigu / acute, s: supérieur / supra, h: haut / high, e: élevé / elevated, m: moyen / mid, c: centré / centred, b: bas / bottom, i: inférieur / infra, g: grave / grave, and I, C, E, S: respectively, first, second, third and fourth F0 level) and the 2 allotones aa and ss.*

Conversely, the ascending tones would be presented in a symmetrical way.

7. Illustration

Figure 3 below presents an example of Praat Textgrid using the MOMEL and INTSMEL procedures.

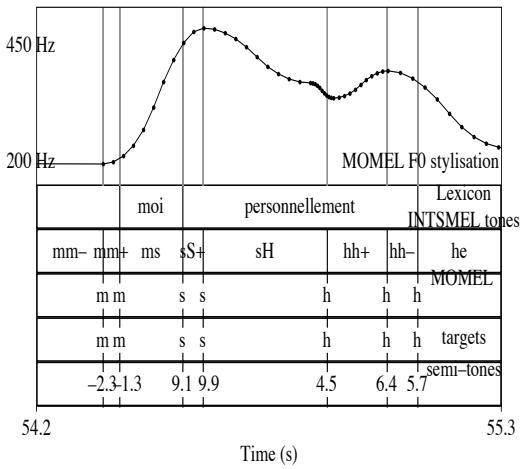

Figure 3. *Example of Praat Textgrid with MOMEL and INTSMEL procedures. For the speech extract ("moi personnellement" / "I, personally") tiers 1, 2, 3, 4, 5 from bottom to top respectively correspond to the semi-tone values, F0 values, the Momel targets, the INTSMEL tones, the linguistic items, and the MOMEL F0 stylisation.*

From bottom to top, tiers 1, 2, 3, 4 and 5 respectively correspond to the semi-tone values, F0 values, the MOMEL targets then the INTSMEL tones, the linguistic items, and the MOMEL-modelled F0 curve.

One can observe that the word *personnellement ("personally")* corresponds to the definition of a melism: indeed, it reaches the *s* level (418 Hz or 9.5 semi-tones above mean for the speaker), it exhibits a 200 Hz or 11.5 semi-tones difference with the F0 level in the previous word *moi ("I")*. This word displays 4 surface phonological tones, respectively sS+, *sH* and *hh+, hh-*. The latter ones result from some minimal F0 variations, all comprised within the limits of the same level *h.*

8. Conclusion

The aim of this paper is to present a surface phonological tonal annotation and stylisation system integrated in the automatic procedure INTSMEL (and MOMEL) and running under Praat. This automatic computation of F0 levels and phonological tones constitutes a particularly useful tool in a specific application exploring the structure and comparing the F0 levels in a local perspective, i.e., in the frame of a word or a sequence of words, and then providing objectivity to this analysis. It can be applied to any language, with lexical stress or not, with tonal system or not. In the near future, some detailed analyses will be presented, using this annotation system in the domain of prosody and pragmatics.

9. References

[1] Caelen-Haumont, G.; Bel, B., 2000. Le caractère spontané dans la parole et le chant improvisés : de la structure intonative au mélisme, *Revue Parole,* 15/16, 251-302, 2000.

[2] d'Alessandro, C.; Mertens, P., 1995. Automatic pitch contour stylization using a model of tonal perception, *Computer Speech and Language,* 9, 257-288.

[3] Di Cristo, A. ; Hirst, D.J., 1986. Modelling French micromelody: analysis and synthesis. In Kholer (ed.): *Prosodic Cues for Segments* (*Phonetica* 43, 1-3), 11-30.

[4] t'Hart, J.; Collier; R., Cohen, A., 1990. *A Perceptual Study of Intonation,* Cambridge University Press, Cambridge.

[5] Hirst, D.; Di Cristo, A., 1998. Intonation Systems: a survey of twenty languages. Cambridge University Press, Cambridge.

[6] Hirst, D.; Di Cristo, A.; Espesser, R., 2000. Levels of Representation and Levels of Analysis for the Description of Intonation Systems. In Horne, M. (ed.): *Prosody : Theory and Experiment. Text, Speech and Language Technbology*, *14.* Kluwer Academic Publishers, 51-87.

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

[8] Pitrelli, J.F.; Beckman M.E.; Hirschberg J., 1994. Evaluation of prosodic transcription labeling relaibility in the ToBI framework, *Proceedings of ICSLP 94,* 123-126.

[9] Malfrère, F.; Dutoit, T.; Mertens, P., 1998. Fully Automatic Prosody Generator for Text-to-Speech Synthesis. *Proceedings of International Conference on Speech and Language Processing.* Sidney, Australia, 1395-1398.

[10] Rossi, M.; Chafcouloff, M., 1972. Les niveaux intonatifs. *travaux de l'Institut de Phonétique,* 1, 167-176.