Fo DECLINATION AS A CUE TO DISCRIMINATION OF TONAL CLASSES AND PHRASING IN FRENCH

Pascal Roméas

Institut de Phonétique d'Aix-en-Provence, France.

ABSTRACT

Fo keypoints follow an overall decay from the beginning to the end of French utterances. This is best accounted for when the keypoints are distributed into 3 tonal classes (L, H, S). We compare the significance of linear and 2nd-order polynomial regressions to account for the Fo declination of these 3 classes. This latter regression generally shows a negative second derivative, which leads to a discussion. We find that class S determines the occurrence of declination resettings. The regression significance may be better, under certain conditions, inside sections delimited by S-points than in the span of the whole utterance. We discuss whether regressions may be a cue to resettings.

1. BACKGROUND.

This study deals with the organisation of the melody keypoints in French utterances, from both frequential and temporal points of view.

The speech material is taken from a French simulated man-machine dialog in which only users' requests have been taken into account.

Our earlier works ([10], [11], [12]) have shown the existence of a two-mode organisation of tone in this type of utterances. We distinguish between :

1-suprasyllabic tonal patterns, whose domain and function refer to the lexical relative information load,

2-intrasyllabic contours, which, along with other redundant cues (pauses, etc.), assume a function at the syntactic level (marking of phrase ends).

These two tonal phenomena can be distinguished from three different points of view : acoustic features, phonological association with the syllable string, functions (at lexical, syntactic, and informative levels).

2. AIM OF THIS STUDY.

Considering that the approaches involving sequences of Fo targets (e.g. [1], [7]) have to be tested on our material, we now examine this tonal organisation in a new manner. The pitch maxima of suprasyllabic patterns have been labelled as H. Those of intrasyllabic contours have been labelled as S. All syllables located off these patterns and contours are considered as unstressed (which for French means : low tone). The center of their vocalic part has been labelled as L. Both time and Fo values of these keypoints have been saved in appropriate files.

Our aim is to show that satisfactory regression functions in the time x Fo space can be found to account for these (x,y) keypoints, under some conditions :

- the functions must be calculated independently for each class of keypoints (H, S, L),
- polynomial functions (generally second order) may often provide a better model than linear regressions,
- the model may often be improved when the utterance has been parsed into sections delimited by the S-points (these sections generally match phrasing, since S-points have a syntactic function).

This paper actually draws the first trends, but complete results and general conclusions will be available in our thesis dissertation by September 91.

3. METHODOLOGY.

We deal with 125 utterances, produced by 5 speakers. Fo calculation and representation, as well as tonal labelling, has been run on a Masscomp-5400 mini-computer, using the SIGNAIX speech signal processor [4]. Data have been transferred to a personal computer in order to run statistics.

The necessity for calculating independent regression functions for H, S, and L, is shown in an indirect way, since it relies on the analysis of variance of the three groups.

For each utterance, and for each tonal class, we compared the R-squared and the probability for linear and for polynomial regressions.

When the utterance had S-points, the same regressions have been tested on the sections delimited by these points. We could then see if the R-squared and p were better in the case of sections.

4. RESULTS.

4.1. Regressions must be applied separately to 3 tonal classes.

We said that acoustic features allow a distinction between tonal events involving H (the so-called suprasyllabic patterns) and tonal events involving S (intrasyllabic contours). The major two features are Fo glide threshold and vowel duration. The average vowel duration in the corpus for all syllables except those bearing an S-point is 85ms. As shown in figure 1, vowels bearing an S-point have much longer durations:

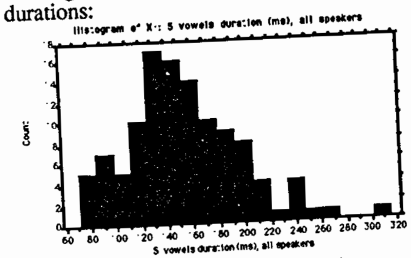

Fig 1: S-vowels duration, all speakers (ms)

mean:	153ms
standard deviation:	43
range:	70 to 304ms
(for 113 values)	

The Fo glide magnitude between S and the preceding L is bigger than the Fo glide magnitude between H and the preceding L:

Ratio Fo(H) / Fo (L), all speakers:

mean Fo(S)/Fo(L):	1.23
standard deviation:	0.11
(for 113 values)	

Ratio Fo(S) / Fo(L), all speakers:

mean Fo(H)/Fo(L):	1.16
standard deviation:	0.09
(for 466 values)	

Otherwise, the regression functions applied to the S group alone have a higher constant than the regression functions applied to the H group alone in 98% of cases.

The analysis of variance of the three groups (H,S,L) confirms that the Fo values organisation in the time dimension must be studied for each group separately. We shall call these groups tonal classes.

4.2. linear vs second order polynomial regressions.

Fo declination is generally described as a progressive Fo downdrift from the beginning to the end of the utterance. Declination models often make a distinction between top-line and base-line downdrift ([1], [8], [9], [13]). As seen in (4.1.), and as shown in figure 2, we find it useful to analyse this phenomenon for 3 separate classes, which provide an L-line, an H-line, and an S-line.

These lines are obtained by regression functions. S-lines have low significance since utterances have few S. However the S-lines slopes do not usually differ significantly from the slopes of other classes. The general shape of the downdrift shows a slight convergence between classes rather than a strict parallelism.

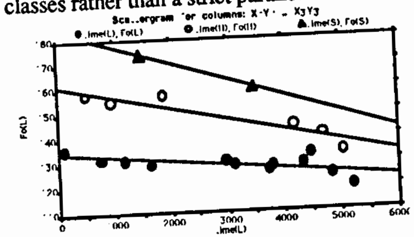

Fig 2: L-line, H-line, and S-line obtained by linear regressions. Time is in milliseconds, Fo in Hertz. Utterance: "j'aimerais connaître le temps prévu le douze juin mille neuf cent quatre vingt deux sur le versant alsacien des Vosges".

The number of H points in the H-lines may be lower than 4 (mean 4.3 per utt.), so that no significant regression can be calculated in these cases (41% of the occurrences). 32% of the H-lines are better accounted for by a linear regression although

approximately 2/3 of these do not reach the probability p=0.1. Actually, if we consider utterances with a greater number of H, the 2nd-order polynomial regression appears to be a better model. This is the case for 27% of H-lines, out of which more than 2/3 have $p<0.05$.
The same tendency can be noticed for L-lines, which gather more keypoints than H-lines (mean 9.6 per utt.). We found that 60% of the L-lines (generally the ones provided with a greater number of L-points) are better accounted for by 2nd-order polynomial functions. Most of them provide a satisfactory R-squared, and over 95% have $p<0.05$.
The x^2 coefficient was found to be negative for over 90% of L-lines. In most cases, the lines can be divided into two temporal phases : first, they increase, but they have a negative second derivative (shorter phase); second, Fo drifts down and the second derivative remains negative (which means that Fo steepens with regard to time). See figure 3.

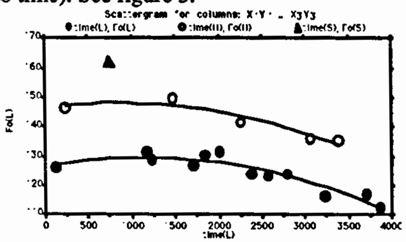

Fig 3: L-line and H-line obtained by a 2nd-order polynomial function. This utterance has only one S-point. R-squared for L is 0.887, $p<0.0001$. R-squared for H is 0.915, $p<0.0855$

A discussion of these results is presented below.
Yet 40% of L-lines are better accounted for by linear regressions. One explanation is that most of these L-lines are provided with few L-points. Moreover, only 1/3 out of those cases reach $p<0.1$ (which is partly due to the low number of points).

4.3. Utterance vs sections.

Trying to find one function to account for keypoints has less and less justification as utterances get longer. Many authors ([1], [3], [13]) have noticed that the course of declination may be reset at major boundaries. Our material provides many long utterances interrupted by silent pauses. The pauses generally occur immediately after S-tones.
We found that roughly half of the L-tones that immediately follow an S-tone (L2) have a higher Fo value than the L-tone that immediately precedes the S (L1). Moreover, if we now consider the Fo difference between L-tones and the y-values of the regression function provided with the same respective x-values, we find that most of the L1 have a negative difference while most of the L2 have a positive difference. This seems to indicate that the resettings must be interpreted with regard to an overall downdrift which covers the whole utterance, rather than to the rough Fo scale. This point deserves further investigation and will not be discussed in this paper.
Another criterion for resetting is the significance of the regression applied to sections delimited by the S-tones, as compared to the regression on the whole utterance.
This criterion is disappointing at first sight. Our hypothesis was that the utterances could be successfully parsed into sections delimited by S-tones. Some utterances do not have S-tones. Otherwise parsing has been attempted as long as S split the utterance in a way which provided the resulting sections with at least one L and one H (thus excluding final S). Finally both linear and polynomial regressions were run on the span of sections. The main problem that we encountered is the lack of points in the sections, especially for H-points.
In cases where the section slopes are obviously reset (relative Fo difference between L2 and L1, silent pause interruptions), the significance of section regressions often remains low. It is lower than the corresponding utterance regressions for 87% of sections, although 38% still provide a p below 0.05.
Yet if we now assume that there is no linguistic reason why the declination models obtained above by regressions on long sequences of points should not be implemented on shorter sequences, we may consider that the R-squared is a better cue than the probability (which is directly related to the degree of freedom). Actually, 63% of section R-squared are higher than the corresponding utterance R-squared.

See illustration in figure 4 & 5.

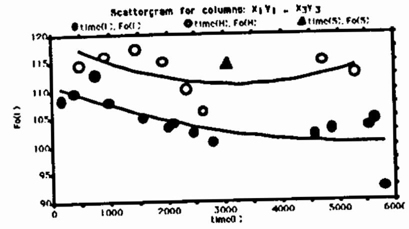

Fig 4: Utterance "Quelles sont les temperatures maxima et minima aux environs de Gérardmer à plus de huit cents mètres aujourd'hui". R-squared for L-line: 0.561, p=0.0108. R-squared for H-line: 0.302, p=0.4076. See next figure for section results.

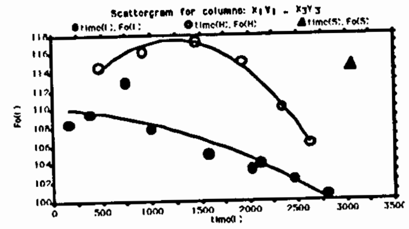

Fig 5: 2nd-order polynomial regressions for the first section of the same utterance as in figure 4. The section ends in the S-point represented by a black triangle. R-squared for L-line: 0.83, p=0.0049. R-squared for H-line: 0.99, p=0.0011.

Further discussions about resetting cues will take place in later publications. We now prefer to focus on the point of the negative second derivative that has been found for most of the declination slopes.

5. DISCUSSION

Negative x^2 coefficient does not confirm the previous observations on the general slope of declination ([1], [5]), which was found to follow an exponential decay (phrase component in Fujisaki's model). This shape of the overall decay has been claimed to be conditioned either by subglottal pressure [6] or by crico-thyroid activity [2]. Beyond this controverse, it may be assumed that declination is mostly determined by the linguistic struture of utterances, and therefore pre-planned independently from physiological constraints ([8], [9], [13]). Since the keypoint values are linguistically conditioned, the exponential decay model cannot be conceived as language- and context-independent. We infer that the steepening slope in our material is due to the specific structure of these French utterances. Since the slope remains a function of time, its shape cannot be conditioned by lexical or syntactic factors. On the contrary, the phatic function may be assumed to weaken smoothly as the S-tone linguistic information nears. The steepening slope may be a pre-indicative cue for the perception of S-tones (i.e. boundaries). As a consequence, the pitch of unstressed syllables in these French utterances should be considered the result of a controlled active process.

REFERENCES.

[1] BRUCE, G., (1984), "Aspects of declination in swedish", *Lund Working papers in Phonetics*, 27, 51-64.
[2]COLLIER, R., (1985), "Setting and resetting of the base-line", *R.I.L.P. Ann. Bull.*, 19, University of Tokyo.
[3] CRYSTAL, D., (1969), *Prosodic systems and intonation in english*, Cambridge University Press.
[4] ESPESSER, R., BALFOURIER, O., (1989), *SIGNAIX, mode d'emploi*, unpublished.
[5] FUJISAKI, H., (1981), "Dynamic characteristics of voice fundamental frequency in speech and singing", *4th symposium F.A.S.E.*, Venice, Apr. 21-24.
[6] GELFER, C., HARRIS, K., COLLIER, R., BAER, T., (1983), "Speculations on the control of fundamental frequency declination", *Haskins Status Report on Speech Research*, 76, 51-63.
[7] HIRST, D.J., (1980), "Un modèle de production de l'intonation", *Travaux de l'Institut de Phonétique d'Aix*, 7, 297-315.
[8] MAEDA, S., (1976), *A characterization of american english intonation*, Ph.D., M.I.T.
[9] PIERREHUMBERT, J., (1979), "The perception of fundamental frequency declination", *JASA*, 66, 2, 363-369.
[10] ROMEAS, P., (1990,1), "Prosodie et lexique : tendances majeures observées en dialogue homme-machine", *Proceedings of the 1st French Congress on Acoustics*, Lyon, April 9-13 1990, Editions de Physique, pp545-548.
[11] ROMEAS, P., (1990,2), "Apport d'information lexicale et marques prosodiques", *18èmes Journées d'Etudes sur la Parole*, Montréal, 28-31 Mai 1990, pp17-20, Pub. de l'Université de Montréal.
[12] ROMEAS, P., (1991), "Organisation prosodique et accès lexical en dialogue homme-machine", *2èmes Journées du PRC-GRECO Communication Homme-Machine*, Toulouse, January 29-30 1991.
[13] SORENSEN, J., COOPER, W., (1980), "Syntactic coding of fundamental frequency in speech production", in COLE, R.(ed.), *Perception and production of fluent speech*, 399-440, LEA, Hillsdale.