SYMBOLIC CODING OF HIGHER-LEVEL CHARACTERISTICS
OF FUNDAMENTAL FREQUENCY CURVES

P. Nicolas and D.J. Hirst
e-mail: {pascal; hirst}@univ-aix.fr
Institut de phonétique d'Aix-en-Provence
Laboratoire "Parole et Langage" URA 261, CNRS
29, Av.R.Schuman, 13621 Aix-en-Provence, FRANCE

ABSTRACT
In this paper we present the results of a study of higher-level
effects on fundamental frequency curves in French. To
observe these, we analysed four readings lasting
approximately one minute of a text composed of three
paragraphs. We used a modelling technique which allowed us
to capture a number of features of individual Intonation Units
embedded in a sentence, in a paragraph and finally in a text.
In our data, we find that the highest intonation structure
seems to be the paragraph and we suggest coding this
structure for TTS systems which generally do not take into
account levels of structure above that of the sentence. Finally,
we propose an extension to the INTSINT transcription
system in order to cover such paratone effects.

1. INTRODUCTION
In a previous study [1] we presented the results of an analysis
of the fundamental frequency structure of continuous texts in
French. We showed that the INTSINT system (International
Transcription System for intonation, [2]) provides a useful
way of modelling the intonation structure of a text. With this
transcription system, we can adequately describe pitch
patterns with a limited set of tonal symbols, each of which
characterises a point on the fundamental frequency curve. The
modelling technique correctly captures a number of features
of individual Intonation Units which are observed in natural
speech, such as declination and relative prominence. It does
not, however, provide any "superordinate structure" above the
level of the Intonation Unit, although such effects have been
described by a number of authors for several different
languages [3, 4, 5, 6, 7]. In these studies, the authors observe
the presence of a "paragraph declination" [6], a
"supradeclination behaviour" [7], a "superordinate F0
declination" [4] throughout the paragraph. This lowering is
generally expressed in the form of a structure defined by the
baseline and the topline. In the majority of these studies, in
order to avoid text-initial and/or text-final effects, only the
effect of sentence position in a paragraph from the middle of
a text, or a text consisting of a single paragraph was
analysed. One of the aims of all these studies, is to find out the
maximum domain of pitch contours.

In prosodic phonology [8], the highest level phonological
unit is the Utterance. Beckman and Pierrehumbert [9]
following Hirschberg and Pierrehumbert [10] found that final
lowering is controlled more by hierarchical discursive
structure, and they conclude that this fact makes the
Utterance incompatible with phonological status. Even if
declination or final lowering is not represented in a
phonological model, speech synthesis systems could be
improved by taking these prosodic parameters into account to
increase the structural consistency of the speech.

A few studies [11, 12, 13] mention the use of the paragraph as
a unit for speech synthesis. These studies essentially deal with
the evolution of pitch range over the sentence and the
paragraph. All these authors base their conclusions on the
evaluation of a version respecting the intonation structure of a
paragraph. This modelling seems not only to improve the
naturalness of the speech synthesis, but also to facilitate the
comprehension of the text. For the moment, this improvement
appears to be fairly slight but significant. The authors
conclude that this fact is due to an incomplete knowledge of
the intonational organisation of the paragraph and to the fact
that no account is taken of the effect of the text level on the
intonation structure.

The purpose of the present study is to present the results of an
analysis of higher-level effects on fundamental frequency
curves in French. In this paper, unlike the studies mentioned,
we take into account an additional level of structure
corresponding to the entire text and we look at the potential
effect on fundamental frequency of the position of the
paragraph in the text.

Finally, we consider the possibility of adapting the INTSINT
system to cover such paragraph effects. The INTSINT
transcription system provides two sorts of tonal symbols:
relative tones (Higher, Lower, Upstepped, Downstepped),
which make reference only to the immediately preceding
pitch-target, and absolute tones (Top Mid, Bottom) which are
assumed to refer to the speaker's overall pitch range within the
current Intonation Unit. We know, however, that this pitch
range is modified when a sentence is integrated into a
paragraph. The question we address is how to model this
change of overall pitch when we examine the distribution of
absolute target points within the sentence, within the
paragraph and finally within the text. How many levels are
necessary to express the structural functions of these absolute
tones?
2. MATERIAL AND PROCEDURE

The corpus consisted of a text composed of three paragraphs for a total of 8 sentences. This text, an extract from an introduction to science for French children, was read aloud by four native speakers of French, two male and two female. The extract has a single topic "the atom". It was presented in normal orthography with its original punctuation. Each reading was recorded in an anechoic chamber and then digitalized at 16kHz on a Sun Sparc station. An automatic modelling of the F0 curve was carried out with a technique called asymmetrical modal quadratic regression \[14\]. The output of this algorithm is a sequence of target points (Hz, ms) defining a quadratic spline function. A small number of erroneous target points were corrected manually. The modelled curve can be used as input for PSOLA-SOLA resynthesis for auditory evaluation of the analysis. In our earlier study \[1\], we presented several different codings of target points some of which require a preliminary segmentation into intonation units while others do not. Here, we use only the first type of coding which obtained a good score in an evaluation of overall acceptability, since criteria used in macro-segmentation are easily automatically reproducible. A macro-segmentation into Intonation Units of our text was carried out using pauses, punctuation and connectors. Once the macro-segmentation was completed, the highest and lowest target points of each unit were coded respectively as T and B, the first point was coded as M (unless already coded as T and B), all other targets were coded with relative tones as either H, L, D, or U according to the value of the preceding and following targets.

3. RESULTS

3.1. Global range

At the text level, there was a slight tendency for lowering of global range within paragraphs in function of the three possible positions in the text, but the differences weren't statistically significant. On the other hand, at the paragraph level, our data (values of T and B) to independent variables like the place of the paragraph in the text, the place of the sentence in the paragraph, the place of the Intonation Unit in the paragraph and in the sentence. We found a difference between the tonal range of the first also tested the effect on T and B of initial, final or medial position in the paragraph, which was greater than that of Intonation Units in final position in sentences. The first effect was stronger than the second. It seems that for tonal range we found two levels of structuring, one at the paragraph level and the other at the level of the sentence. There were thus two levels of structuring corresponding to three observational levels: the paragraph, the sentence, and the Intonation Unit. Three phonetic parameters were analysed: tonal range expressed by the range at the paragraph level and the other at the level of the between target points coded T and B within the intonation unit. We can notice a considerable variability in lowering of the target point T and lowering of the target point B throughout the text.
and a min within an IU is not the best way to express the sequential lowering within the text and its subunits. We found better results and less data variability by analysing separately two sorts of targets points T and B which don’t seem to have the same function in the intonational structure of the text.

3.2. T target point lowering through the paragraph

At the text level, the only significant difference found was between the mean of the Ts of the first and the last sentences of the text (around 10% mean difference). However this difference was no longer significant when measured on an ERB scale [15].

The pitch range of target points coded T seems to be used to indicate the beginning of paragraphs. Thus, we found a significant result (p=0.0153) for the difference between the T of the first IU and the other Ts of the paragraph. In this case, means of T decrease for all speakers in the direction initial, medial and final IU positions within the paragraph.

But this type of target point also serves to mark the start of a sentence. There was a significant difference (p=0.0025) between the value of the T of the first IU of sentences compared to that of the last IU of sentences (around 9.5% increase). In the few cases where a sentence corresponded to a single IU we found almost the same mean value for T as for those in the first IU of a sentence which seems to indicate, contrary to what has sometimes been suggested [16, 3] , that pitch onset is not correlated to the length of the following sentence (for similar conclusions [4]).

In our data, although we found a hierarchical lowering of T in function of the position of the IU in the paragraph, we didn’t find a significant difference between the height of T in sentence-initial and paragraph-initial position. In fact sometimes, in our data, sentence-initial Ts have higher values than paragraph-initial Ts.

Another unexpected result was the contribution of the value of T to marking the end of a paragraph (p=0.0171). The last IU of a paragraph seems to get a global lowering stronger than the others. In this way, the perceptual mark of paragraph finality occurs in quite a large unit which suggests that a listener can detect the end of the paragraph well before the end.

3.3. B target point lowering through the paragraph

We observe a mean lowering of about 12.5% of the value of B in the last IU in the paragraph (p=0.0059). We found the same result for the relation IU to sentence where we found a significant difference between the last IU of the sentence and the others (p=0.0005). On the other hand, the difference between the value of B in initial and final IUs within the paragraph was not significant. Unlike the target point T which seems to contribute to the marking of both the beginning and the end of a paragraph, the target point B seems only to play a role in marking finality.

We found a significant difference between the value of B in the last IU in a sentence/paragraph and that in an IU within the sentence (p=0.0001). Once more, there was no significant difference between target points B located at the end of a sentence and those at the end of a paragraph although there was a tendency for the means of B to decrease for all speakers in the direction end of IU, end of sentence and end of paragraph.

Even though this difference is not significant, we can note that the mean of B belonging to the first IU of paragraph is a bit lower than the mean obtained with the target point B of the sentence initial IU.

4. DISCUSSION

In our data, we found no major significant effects related to the level of text subdivided into paragraphs. The maximal super-ordinate intonation structure as regards to lowering seems to be the paragraph (or paratone) rather than the whole text. On the other hand, our results confirm those of earlier studies on the role of the paragraph. We see that tonal lowering emerges more clearly from an analysis of the isolated target points T and B than from a measurement of the tonal range of a macro-unit. Target points T and B have different functions to mark the structure within sentences and within paragraphs.

Concerning the introduction of a possible paragraph level in a transcription system, we saw that the paragraph is above all marked for finality. When we look at definitions of the paratone [17], we see that it is usually attributed two characteristics: it begins with a high onset (or high key), it ends with very low pitch and the presence of a noticeable pause. While the second characteristic was verified in this
study, the first one was not completely verified, but we shall of course need to examine many other texts to confirm these results.

One aim of this study is to present and discuss ways in which this higher-level structure may be incorporated into an extended version of the INTSINT transcription system. Symbols used in INSINT are of two types: boundary symbols and tonal symbols. In figure XXX, corresponding to an example of the phonetic evolution of two absolute target points T and B throughout the paragraph, we have introduced diacritic symbols (+ or -) to express the observed significant differences in tonal range of target points T and B throughout the paragraph. Such diacritic symbols would not in themselves be a satisfactory extension of INTSINT. If the symbols were freely available that would suggest that any sequence of T, T+, T- or B, B+ and B- would be a possible sequence. In fact they always lower so that within a given sentence the sequence goes from T+/B+ to T/B to T-/B- etc. It seems that a more adequate solution would be a phonological marking which would account for both the sentence-level and the paragraph-level effects. There is, however, an apparent paradox here since while both sentence and paragraph level effects seem to be relevant there was no difference observed between the onset (T) values of sentences and paragraphs, nor between the offset values (B) of sentences and paragraphs. There was, however a significant effect observed on the onset of the final IU of paragraphs. Hirst (in press), suggested for British English that higher order effects could be accounted for by using double brackets to delimit a higher order prosodic unit representing the paratone. The effects we have observed could be accounted for by using double brackets for the beginning of sentences and the end of paragraphs. The example above would thus be represented:

\[
[[[T B][T B][T B][T B][T B][T B][T B][T B][T B]]]
\]

Rules of phonetic interpretation would then apply to this structure to derive the observed patterns of relative height.

It remains of course to be seen whether this solution can be generalised to other data.

---

**REFERENCES**


