

Characteristics of the Accentual Phrase in French: an Acoustic, Articulatory and Perceptual Study

Guillaume Rolland and H  l  ne L  evenbruck

Institut de la Communication Parl  e
INPG – Universit   Stendhal – CNRS UMR 5009
Tel: +33 4 76 57 47 14 - Fax: +33 4 76 57 47 10
e-mail: {rolland; loeven}@icp.inpg.fr

Abstract

The purpose of this study is to evaluate the acoustic, articulatory and perceptual relevance of the Accentual Phrase (AP) in French. The AP is the lowest level of the hierarchical model of French intonation proposed by Jun and Fougeron [7]. A corpus of 22 sentences, each likely to yield 3 APs, was recorded for 4 native French speakers. Our first aim was to describe the acoustic properties of the AP. We measured both primary (LH*) and secondary (LHi) accents in terms of F0 peak values, F0 slopes, duration of the F0 rises, duration of the peak-bearing syllables and number of syllables needed for the rise. We found that the AP was often realized as [LHiLH*], with the H* peak featuring a higher F0 value, a longer duration, a stronger syllable-anchoring than Hi, and the LH* rise being more dynamic than LHi. We compared these acoustic results with those of a previous articulatory analysis and confirmed the preponderance of H* over Hi. Then using perceptual tests, we tried to evaluate the ability of French listeners to slice utterances into Accentual Phrases. The slicing scores show that the listeners could well split the sentences into separate APs. Once this ability quantified, we tried to match these results with the acoustic measurements to address the following question: what makes an AP be perceived as a unit? We suggest that the Accentual Phrase is perceived as a unit mainly because of its tonal (low-high-low-higher), durational (AP-final syllable longest) and dynamic (long small rise – brief large rise) characteristics.

1. Introduction

1.1. Goal

If the production (acoustic, articulatory) aspects of prosody are largely studied, the perception correlates are often neglected. However like other components of phonology, the prosodic structure depends on physiological *and* perceptual constraints. The purpose of this study is to link production and perception characteristics of the Accentual Phrase (henceforth AP), the lowest unit of a hierarchical model of French prosody.

1.2. Context

Many phonological models of the prosodic structure of French have been proposed (e.g. [13], [16], [3], [4], [10], [12], [5]). These models consider in common that French intonation is a sequence of rising pitch movements, that accent is postlexical and that utterances are organized into a

hierarchy of different prosodic levels. But the precise hierarchy, the number of levels, the tonal and intonational default patterns are still discussed, even the notion of accent in French is debated.

Jun & Fougeron's model, which follows the framework developed by [11], [14] and [1], agrees with most descriptions of French prosody, and uses a transcription similar to the prosodic transcription tool ToBI (Tones and Break Indices) [15], which principles are used in many languages. This model features two prosodic units: the Accentual Phrase (AP) and the Intonational Phrase (IP) which correspond respectively to the low and high levels of the hierarchy (which follows the Strict Layer Hypothesis).

The lowest hierarchy level, the AP, contains one or more content words and is right-demarcated by what has been called the primary stress (H*). An initial LH (Low-High) tonal sequence can mark the initial boundary of an AP. This LHi sequence corresponds to what has been called the initial or secondary accent. The default tonal pattern of the AP is thus /LHiLH*/.

Higher in the hierarchy is the IP level, which can preempt the AP level. If an AP is IP-final, the default H* tone will be replaced by the boundary tone of the IP (L% or H%, see [7]).

2. Production: acoustic and articulatory correlates of the AP

The purpose of this part is to look at the physical manifestations of the AP in speech production.

2.1. Acoustic characteristics of the AP

In the acoustic domain, our aim is to check whether the tonal patterns of declarative sentences match with the predictions given by Jun & Fougeron's model.

2.1.1. Corpus

A corpus of 22 sentences with a Subject – Verb – Object syntactic structure (SVO) and with CV syllables (voiced consonant – vowel) has been built. Each of these sentences could *a priori* be considered as a single IP consisting of 3 APs. The number of syllables in the Subject and Object APs could vary from 3 to 5 and in the Verb AP, from 1 to 5. Sixteen sentences had a balanced structure (equal number of syllables in each syntactic phrase) and 6 had an unbalanced structure, as exemplified in the samples below, respectively a) and b):

a) balanced sentences

3-3-3: [Mon mari]_S [ranima]_V [le marin.]_O

My husband revived the sailor.

4-4-4: [Le rat marron]_S [voulait manger]_V [le long mulot.]_O

The brown rat wanted to eat the long field mouse.

5-5-5: [Le marin roumain]_S [voulait ranimer]_V [la jolie maman.]_O

The Romanian sailor wanted to revive the pretty mother.

b) unbalanced sentences

3-1-6: [Le marin]_S [vend]_V [les longs navets jaunis.]_O

The sailor sells the long yellowed turnips.

4-1-2: [Le long mulot]_S [vend]_V [le rat.]_O

The long field mouse sells the rat.

5-1-2: [Le mauvais marin]_S [vend]_V [le rat.]_O

The bad sailor sells the rat.

The 22 sentences have been read aloud by 4 native French speakers (2 female and 2 male speakers). 88 utterances were thus available for analysis. The recording took place in a sound-proof room using the EUROPEC software. Acoustic signals were sampled at 16kHz.

2.1.2. Experimental measurements

The imposed segmental constraints (the use of voiced phonemes and CV syllables) facilitated the pitch tracking and made the prosodic transcription easier. The transcription has been carried out by three experts (well-trained listeners) with a similar notation as the one used for ToBI, as shown in the example given in Figure 1. Pitch peaks were labeled and their F0 values measured. Syllable durations, pitch rising durations (in milliseconds and also in number of syllables), pitch rising slopes (Hz/s) were calculated. We were then able to compare the realizations of Hi vs. H* in terms of duration, number of carrying syllables, F0 rises (movements and reached values) as well as in terms of intensity (not described in our analysis). Finally, we could compare the AP tonal realizations with the predictions of the model.

2.1.3. Results

According to the model, the default tonal pattern of our 22 sentences, which *a priori* contained 1 IP made up of 3 APs (one for the Subject, one for the Verb Phrase, and one for the Object) was /LHiLH* LHiLH* LHiL%/. We note however

that, according to the model, the actual tonal realization of an AP depends on several factors, such as the number of syllables, the speech rate or style, phonological constraints, and rhythmic principles. APs with an insufficient number of syllables (less than 4) can be realized [LH*], [LLH*], [LHiH*] or [HiLH*].

In the present study, the acoustic analysis shows that 69% of the initial APs (initial position in the utterance, Subject) were realized as [LHiLH*]. When their number of syllable was at least equal to four, 91% of the initial APs displayed the default tonal pattern. The central AP (central position in the utterance, Verb Phrase) was realized [LHiLH*] in only 47% of the cases, but recall that the number of syllable varied from 1 to 5 in this position. Indeed, when the number of syllable reached at least four, 70% of the central APs displayed the default tonal pattern.

When Hi and H* were both realized, we quantified and compared their variations. It appears that the F0 value reached at the peak is significantly higher for H* than for Hi in both initial and central APs (respectively $F(1,57)=14.05$; $p<.05$ and $F(1,30)=23.31$; $p<.05$). The syllable which carries the peak is significantly longer for H* than for Hi in both initial and central APs ($F(1,57)=149.62$; $p<.05$ and $F(1,30)=11.85$; $p<.05$). The difference in the slope of the F0 rise is not significant ($p>.05$). The number of syllables needed for the F0 rise is significantly higher for Hi than for H* in both initial- and central-AP ($F(1,57)=60.69$; $p<.05$ and $F(1,30)=29.21$; $p<.05$). We also noticed that the H* pitch rise is carried by the last syllable in 93% of the initial APs, and in 88% of the central APs. Finally the duration of the F0 rise (in ms) is significantly higher for Hi than for H* in initial APs ($F(1,57)=27.83$; $p<.05$).

In summary, for an initial AP, the primary accent features a higher pitch peak, and a shorter rise which also takes less syllables than Hi. The syllable that carries the pitch peak is longer than the one for Hi. The primary accent is usually entirely carried by the AP-final syllable only (rise included) whereas the initial rise towards the Hi peak can be carried by 2 or 3 syllables. When the AP is central in the utterance, the conditions are the same except for the duration of the rise ($p>.05$).

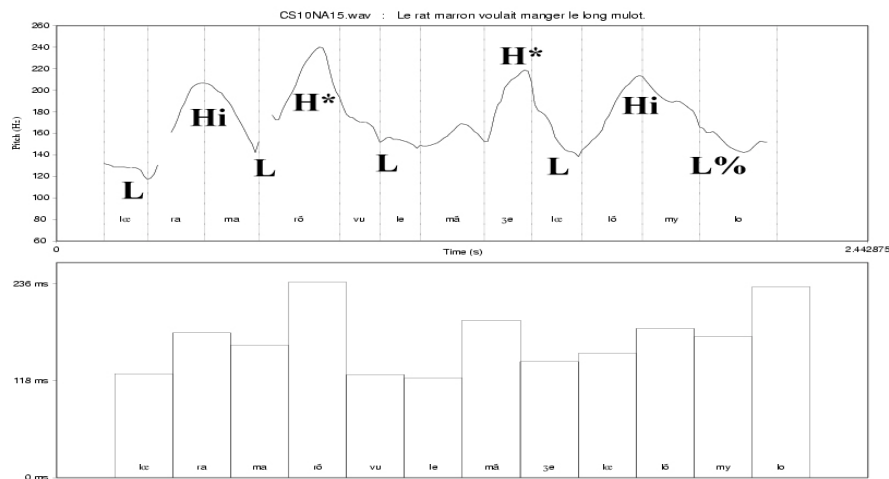


Figure 1: Pitch tracking and syllable durations for {[Le rat marron]_{AP1} [voulait manger]_{AP2} [le long mulot.]_{AP3}}_{IS} (The brown rat wanted to eat the long field mouse.). The corresponding transcription is {[LHiLH*] [LH*] [LHiL%]}.

2.1.4. Preliminary conclusion

The acoustic analysis revealed a difference between the primary and secondary accents in French. Our results suggest that H* is probably perceived as the primary accent because it reaches a higher F0 value, and because the F0 rise is also more dynamic than that of Hi (larger in amplitude and shorter in time) and born by a single syllable. This relates to Lehiste [8] who showed that a syllable is perceived as more accented when surrounding syllables are unaccented (flat F0 and no lengthening). The aim of the following part is to compare these results with those of an articulatory study.

2.2. Articulatory characteristics of the AP

2.2.1. Results of a previous articulatory study

In a previous articulatory analysis [9] using a magnetometer (Carstens EMA), tongue movement durations and amplitudes have been compared for the initial (Hi) and final (H*) accents, and for two speaking modes: natural and under contrastive focus. Acoustic and articulatory signals have been recorded for 2 speakers (a female and a male).

2.2.2. Preliminary conclusion

In the natural mode, it appears that Hi was, in most of the cases, weaker than H* in terms of duration and movement amplitude. It reinforces the idea that initial and final accents have a different status.

The influence of contrastive focus was variable according to the speaker. For the male speaker, contrastive focus had little effect and only seemed to slightly reinforce the strength of H* vs. Hi. For the female speaker, contrastive focus had more impact, lending Hi a strength similar to that of H*.

2.3. Conclusion : physiological characteristics of the AP

Our acoustic and articulatory results corroborate the assumption that the primary accent is stronger than the secondary one. The AP can be described as featuring a /LHiLH*/ default pattern, with the F0 value of the peak, the amplitude of the tongue movement and the duration of the peak-bearing syllable higher for H*. In order to verify that the AP is indeed perceived as a unit by listeners, we carried out a perceptual analysis, described below.

3. Perception: perceptual correlates of the AP

Is the lowest level of the hierarchical model perceived as a single unit? In that case, are lexical and syntactic contents the only parameters used to slice utterances into accentual units? To answer these questions, perceptual tests with read- and reiterated-speech have been carried out. Diphone concatenation synthesis using the PSOLA algorithm has been used to build the reiterated speech stimuli from the original recordings.

3.1.1. Experimental method

The 88 sentences recorded for the acoustic analysis have been used again for the perceptual task. Listeners were asked to pick, among 5 choices, which slicing they found most adequate, by paying attention to melodic cues. They could

listen to the stimuli 3 times before selecting one of the choices, as shown in the example below, for the utterance “[Le long mulot]_S [voulait manger]_V [le rat marron.]_O” ([The long field mouse]_S [wanted to eat]_V [the brown rat.]_O):

- 1 – (Le long mulot voulait manger le rat marron.)
- 2 – (Le long mulot voulait manger) (le rat marron.)
- 3 – (Le long mulot) (voulait manger le rat marron.)
- 4 – (Le long mulot voulait) (manger le rat marron.)
- 5 – (Le long mulot) (voulait manger) (le rat marron.)

For each listener, two tests were carried out: one using the original corpus (as in the previous example) and a second one with reiterated speech (delexicalized corpus), the syllable “ma” replacing all syllables. Each of the 88 sentences in each corpus was presented three times, in a random order. Ten native French speakers but naïve listeners listened to the stimuli through headphones and answered the 528 questions (88x3x2), using a program and a GUI (Graphic User Interface) developed under MATLAB. 8 subjects were tested on the original corpus first, 2 on the delexicalized corpus first.

3.1.2. Results

The difficulty of the task explains the results presented in figure 2, which shows the variance found in the percentages of good answers given by the listeners. An answer is considered as correct when it corresponds to the answer given by experts who were helped by the prosodic transcription (see 2.1). The correct answers mainly correspond to #5 (75%) and #3 (20%) slicing choice, and sometimes #4 (5%). Neither the choices number 1 and 2 were rated as correct in more than 1% of the cases.

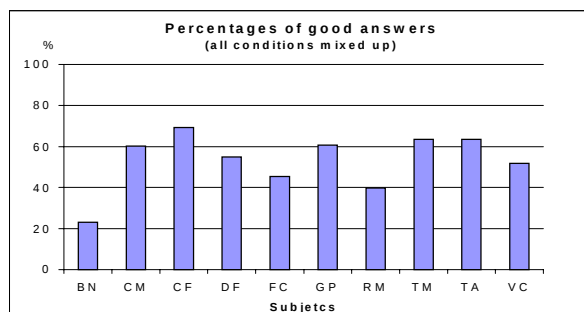


Figure 2: Results of the perceptual tests, all conditions mixed up.

The average score (whatever the corpus and whatever the sentence type) of 53.2% is high enough, compared to the chance level (20%), to consider that French listeners managed to slice utterances into APs. We would like to notice that, when the answer was not correct, listeners often selected a choice with a boundary between Subject and Verb, which makes us confirm Vaissière’s observations [16] on the strength of this boundary.

To figure out whether lexical and syntactic contents constituted the dominant information used in the slicing task, we compared the scores obtained in the original corpus with those of the delexicalized corpus. Figure 3 presents the scores of all listeners for each corpus (lexicalized vs. delexicalized).

The statistical analysis shows that the difference between the scores for each corpus (lexicalized vs. delexicalized) is not significant ($p > .05$): the complexities of the two tasks can be considered as equivalent. These results imply that:

- French listeners do use prosodic cues to segment utterances, since this is the only information available in the delexicalized corpus.
- Lexical and syntactic contents are not the primary information used to recover prosodic phrasing (the average score for the delexicalized corpus was actually slightly, although not significantly, higher than that for the lexicalized corpus).

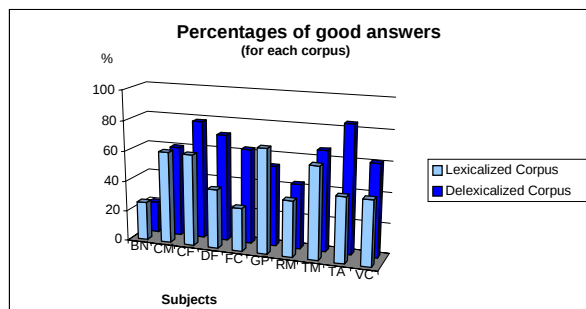


Figure 3: Results of the perceptual tests, for the lexicalized and delexicalized corpora.

A further statistical analysis concerned the comparison of the scores obtained in the balanced vs. unbalanced sentences. The difference between the scores obtained for the balanced- and unbalanced-structured sentences is not significant: the complexity of the slicing task is the same whatever the syllabic repartition of the utterances.

We then tried to characterize which APs were best perceived by the listeners. We recall that, according to the model, the first 2 APs of each sentences possessed the default /LHiLH*/ pattern, but could effectively be realized differently, depending on the number of syllables and on other constraints. Our aim was to answer the following question: what makes an AP be perceived as a single unit? We kept the sentences whose scores were higher than 70%. This means that for a given sentence, the expected answer was selected at least 21 times over 30. According to our results, 78% of these well-perceived sentences feature at least one [LHiLH*] pattern among the first 2 APs (88% for the lexicalized corpus vs. 72% for the delexicalized one). This score is high enough to let us think that the [LHiLH*] tonal pattern facilitates the perception of an AP as a single unit.

3.1.3. Conclusion of the perceptual study

Our perceptual tests bring us to the conclusion that the AP is indeed perceived as a single unit. With no lexical information, listeners are still able to properly slice utterances into accentual phrases, with the only use of prosodic parameters. Moreover, they do obtain a very high score when the tonal pattern of the APs is the default [LHiLH*].

4. Conclusion

These first results seem to validate the acoustic, articulatory and perceptual relevance of the AP, the lowest unit of Jun and

Fougeron's hierarchical model of French intonation. We expect to extend our study to audiovisual perception. If prosodic cues do exist for the audio component, we are free to think that they also exist for the visual one. Recent works have shown a visual coding for Cantonese tones [2].

5. Acknowledgements

We would like to thank our 6 speakers and the 10 listeners for their participation, as well as Marie-Agnès Cathiard, Angélique Grosgeorges and Jean-Luc Schwartz for their help in the development and the analysis of the perception tests, and Cécile Fougeron for her help, comments and advice.

6. References

- [1] Beckman M.E. & Pierrehumbert J., 1986. Intonational structure in Japanese and English. *Phonology Yearbook*, 3, 255-309.
- [2] Burnham D., 2001. Visual discrimination of Cantonese tones by tonal but non-Cantonese speakers and by non-tonal language speakers. *AVSP 2001*, Aalborg, Denmark, 155-160.
- [3] Di Cristo A. & Hirst D., 1993. Rythme syllabique, rythme mélodique et représentation hiérarchique de la prosodie du français. *Travaux de l'Institut de Phonétique d'Aix-en-Provence*, 9-24.
- [4] Di Cristo A., 1998. Intonation in French. In *Intonation systems: a survey of twenty languages*. Hirst D. & Di Cristo A. (eds.). CUP, 195-218.
- [5] Jun S.-A. & Fougeron C., 1995. The Accentual Phrase and the prosodic structure of French. *Actes de ICPhS 95*, Stockholm, 2, 722-725.
- [6] Jun S.-A. & Fougeron C., 2000. A phonological model of French intonation. In A. Botinis (ed.) *Intonation: Analysis, modeling and technology*. Dordrecht: KAP, 209-242.
- [7] Jun S.-A. & Fougeron C., in press. Realizations of Accentual Phrases in French intonation.
- [8] Lehiste I., 1970. *Suprasegmentals*. Cambridge, MIT Press.
- [9] Løevenbruck H., 2000. Effets articulatoires de l'emphase contrastive sur la Phrase Accentuelle en français. *Actes des XXIIIèmes JEP*, Aussois, France, 165-168.
- [10] Mertens P., 1993. Intonational grouping, boundaries and syntactic structure in French. *Proceedings of the ESCA Workshop on Prosody*, Lund, 41, 155-159.
- [11] Pierrehumbert J., 1980. *The Phonology and Phonetics of English Intonation*. Ph.D. dissertation, MIT.
- [12] Post B., 1993. *A Phonological Analysis of French Intonation*, MA thesis. Univ. Nijmegen.
- [13] Rossi M., 1985. L'intonation et l'organisation de l'énoncé. *Phonetica*, 42, 135-153.
- [14] Selkirk E., 1984. *Phonology and Syntax: the relation between sound and structure*, Cambridge, Mass.: MIT Press.
- [15] Silverman K., Beckman M. E., Pitrelli J., Ostendorf M., Wightman C., Price P., Pierrehumbert J., Hirschberg J., 1992. ToBI: A standard for labelling English prosody, *ICSLP 92*, 867-869.
- [16] Vaissière J., 1997. Langues, prosodies et syntaxe, *A.T.A.L.A.*, 38 (1), 53-82.