ISCA Archive
http://www.isca-speech.org/archive

Speech Prosody 2004
Nara, Japan
March 23-26, 2004

Identification of the Possible Visible Correlates of Contrastive Focus in French

Marion Dohen, Hélène Lœvenbruck, Marie-Agnès Cathiard & Jean-Luc Schwartz

Institut de la Communication Parlée
UMR CNRS 5009, INPG, Stendhal Univ., Grenoble, France
{dohen; loeven}@icp.inpg.fr

10.21437/SpeechProsody.2004-18

Abstract

This study aims at determining whether there are visual cues to contrastive focus in French. An audiovisual corpus was recorded from a male native speaker of French consisting of sentences with a subject-verb-object (SVO) syntactic structure. Four conditions were studied: focus on each phrase (S,V,O) and broad focus. The corpus was first acoustically validated: the pitch maximum over the utterance was generally on a focused syllable and duration and intensity were higher for the focused syllables. Then lip area and jaw opening were extracted from the video. The analysis of the data enabled us to extract a set of visible correlates of contrastive focus in French: a) increase in lip area and jaw opening on the focused item b) lengthening of the prefocal syllable and of the focal syllables (even more significant on the first segment of the focused phrase). Thus, there are visual cues to contrastive focus that may be used in communication.

1. Introduction

1.1. Why study “visual” prosody?

Contrastive focus is used to emphasize a word or group of words in an utterance as opposite to another. In French, it can be either syntactic (“C’est x qui a mange la pomme.” *It was x who ate the apple.*) or prosodic (“Wf a mange la pomme.” *Xf ate the apple*.). We will study prosodic contrastive focus here and will use the phrase “S (V or O) focus” in meaning prosodic contrastive focus on S (V or O).

Studies of French prosody have mainly focused on laryngeal and pulmonic correlates of prosody. A few supralaryngeal analyses exist, mostly considering tongue movements [1] or spectral consequences of differences in articulation [2]. Very few studies have examined visual cues to prosody. Those who have done so have focused on visible consequences of F0 variations [3] or on facial cues such as eyebrow movements [4]. Only few studies have examined visible mouth correlates ([5,6,7]) and none for French. “Visible” mouth correlates include visible articulatory correlates such as mouth opening and durational ones, such as syllable lengthening. The purpose of this study is to relate tonal and visual characteristics of contrastive focus in French.

1.2. Background

Jun & Fougeron’s model [8,9] was used in the present study. It agrees with most descriptions of French intonation and uses a transcription system consistent with the widely used ToBI [10]. It features two hierarchical prosodic units. The lower is the Accentual Phrase (AP, right demarcated by the primary stress (H*) and sometimes marked by an initial LHi (Low-High) tonal sequence called the secondary accent). The default tonal pattern of the AP is /LHiLH*/ as realized on the second AP of Figure1a). The higher prosodic unit is the Intonational Phrase (IP) which can preempt the AP level. E.g., if an AP is IP-final, H* is replaced by the boundary tone of the IP (L% or H%) as shown in the last AP of Figure 1a).

In this model, contrastive focus is considered to be marked by a strong Hf and by a low plateau on the subsequent syllables. Hf most often replaces Hi (Figure 1b), but it can also replace both Hi and H* (i.e. the rise in F0 is carried by all the syllables in the phrase and culminates on the last syllable).

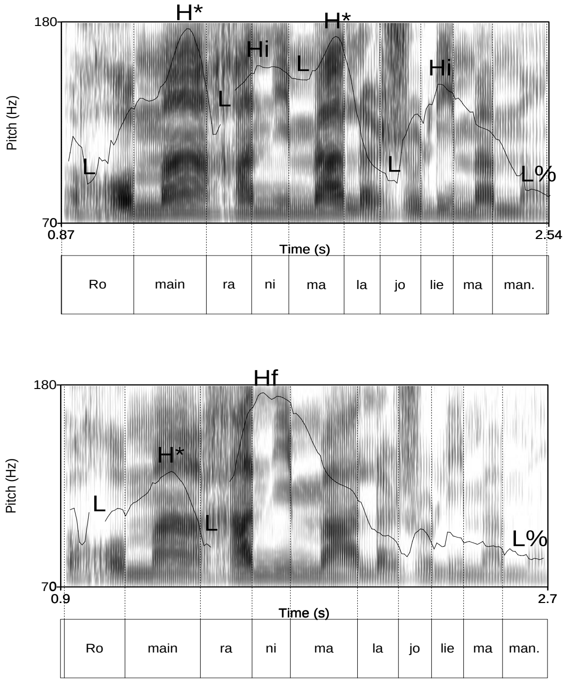

Figure 1: *spectrogram and F0 trace for an IP including 3 APs. a. (top) broad focused case. b. (bottom) focus on the verb AP. The utterance was* $\{[\text{Romain}]_{AP}[\text{ranima}]_{AP}[\text{la jolie maman}]_{AP}\}_{IP}$ *(Romain revived the pretty mother.).*

2. Experimental method

2.1. The corpus

The corpus consisted of eight sentences with a Subject-Verb-Object syntactic structure (SVO) and with CV syllables. Each sentence was likely to be produced as a single IP consisting of 3 APs. In the broad focus condition, the default tonal pattern is thus expected to be {[LHiLH*]$_S$ [LHiLH*]$_V$ [LHiLL%]$_O$}. When possible, we favoured sonorants in order to facilitate the F0 tracking. Below are the eight sentences used.

s1.[Jean]$_{S1}$ [veut ménager]$_{V3}$ [nos jolis nouveaux navets]$_{O7}$.
'Jean wants to spare our fine new turnips.'
s2. [Romain]$_{S2}$ [ranima]$_{V3}$ [la jolie maman]$_{O5}$.
'Romain revived the good-looking mother.'
s3. [Mélanie]$_{S3}$ [vit]$_{V1}$ [les mauvais loups malheureux]$_{O7}$.
'Melanie saw the unhappy bad wolves.'
s4. [Véroniqua]$_{S3}$ [mangeait]$_{V2}$ [les mauvais melons]$_{O5}$.
'Veronica was eating the bad melons.'
s5. [Les mauvais loups]$_{S4}$ [mangeront]$_{V3}$ [Jean]$_{O1}$.
'The bad wolves will eat John.'
s6.[Mon mari]$_{S3}$ [veut ranimer]$_{V4}$ [Romain]$_{O2}$.
'My husband wants to revive Romain.'
s7.[Les loups]$_{S2}$ [suivaient]$_{V2}$ [Marilou]$_{O3}$.
'The wolves were following Marilou.'
s8.[Le beau marin]$_{S4}$ [vit]$_{V1}$ [Véroniqua]$_{O4}$.
'The good-looking sailor saw Veronica.'

2.2. The audio-visual recording

The corpus was recorded from a male native speaker of French with front and profile cameras (see Figure 2). Four conditions were elicited: subject-, verb- and object- focus (narrow focus) and broad focus (broad focus). In order to trigger focus, the speaker had to perform a correction task by focusing a phrase which had been mispronounced in the prompt. The recording went as follows (where capital letters signal focus):

Audio prompt: Denis ranima la jolie maman.
Speaker uttered: ROMAIN ranima la jolie maman.

The speaker was given no indication on how to produce focus (e.g. which syllables should be accented). Four speaking modes were recorded: real, reiterant speech, whispered and reiterant whisper. Yet, only two modes have been studied until now: real and reiterant. Reiterant speech was produced by replacing all the syllables with [ma]. The purpose of reiterant speech is to be able to compare the acoustic and articulatory features across all the syllables. 256 utterances were recorded (8 sentences, 4 focus conditions, 4 speaking modes, all were recorded twice) and 128 have been studied.

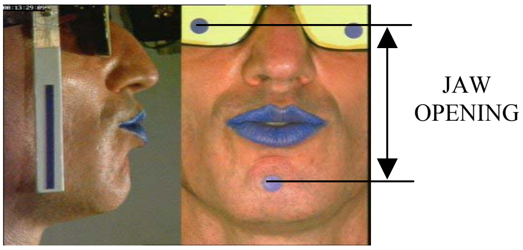

Figure 2: *Video signal recorded: measurement method.*

2.3. Tonal validation of the corpus

This preliminary study aimed at confirming that the speaker had pronounced the focused phrases with a typical focus intonation. For each production, it was checked that the F0 maximum over the whole utterance was on one of the focused syllables. When it was not, we carefully listened to the utterance and verified that it was due to declination. A focused object phrase (utterance-final) may actually display F0 peaks of equal (or even smaller) magnitude to those of the subject phrase (utterance-initial). Declination is however known to be compensated for by listeners [11]. It was also checked that F0 was higher on the focused syllables. We verified that the first content word syllable of the focused phrase carried a Hf accent, as described in [8]. We showed that, in average, the F0 maximum in a phrase was higher when it was focused. Taking declination into account, F0 was always higher on the focused phrase. Similar conclusions were drawn from intensity. We also measured a rise in the mean duration of the focused syllables. The measurement of the mean duration of the post focal syllables showed no significant change from the unfocused to the focused utterances. As described in other studies (see e.g. [8]) we found a deaccentuation of the post focal sequence but not a dephrasing (the phrasing information is cued by the duration information). The items therefore clearly contained cues to focus structure consistent with previous observations [8,12,13,14].

2.4. Measurement techniques

Figure 2 shows an example of the images that were recorded. A program designed at Institut de la Communication Parlée (ICP) [15,16] enabled us to extract parameters describing lip shape and protrusion and jaw position from a sequence of digitalized frames. The mouth opening gesture was studied through a marker on the jaw (see Figure 2). The lip contour was detected from the video signal and lip height, spreading, area and protrusion were extracted from this contour.

3. Preliminary study: reiterant speech

Before studying real speech, a preliminary study was conducted on reiterant speech [17,18]. The purpose was to determine a set of possible visible correlates to contrastive focus. These results showed that the **large jaw opening** gestures associated with **high opening velocities** on all the focused syllables and the **long lip closure for the first segment** of the focused group could be interpreted as a set of visual cues to the perception of focused reiterated [ma] sequences. Additional cues may be **prefocal lengthening** and **post-focal hypo-articulation**.

4. Analysis of the potential visible correlates of contrastive focus in French for real speech

4.1. Preliminary analysis of the problem

Two kinds of visible correlates must be taken into account: articulatory movements and durational variations. The parameters measured (lip height and jaw opening) in the preliminary study for reiterant speech were chosen because their variations distinguished focused and unfocused conditions for the syllable considered ([ma]). For real speech however, all the syllables are different, thus articulatory parameters potentially significant for all the syllables are needed. Recall that unlike for reiterant speech, the articulatory parameters for real speech not only vary because of prosodic changes but also because the syllables are articulatorily different (e.g. jaw and lip opening or duration are not the same from /ga/ to /mi/).

4.1.1. Possible articulatory correlates

There are many possible visible articulatory correlates: jaw opening, lip- height, area, spreading, protrusion, etc. The problem is to identify the one(s) which will vary the most significantly across conditions and the most invariantly across syllables. In our preliminary study [17] we found that the main articulatory consequence of contrastive focus is hyper-articulation. Hyper-articulation can be achieved in various ways, including increase in the amplitude of lip and/or jaw opening and closing movements, increase in lip spreading or narrowing. The parameter affected by hyper-articulation varies, depending on which syllable is uttered: for a hyper-articulated /a/ the mouth will be more opened thus the lip opening and the jaw opening will therefore be larger, for a hyper-articulated /i/, lip spreading, but not lip height will increase, and for a hyper-articulated /u/, lip protrusion will increase but not lip height nor spreading. Taking into account that our corpus contains only very few syllables for which protrusion could be affected by hyper-articulation we did not study this parameter. The parameters which are most likely to be affected by hyper-articulation in this corpus are thus lip height (LH), lip spreading (LS) and jaw opening (JO). The lip area parameter (LA) takes into account the variations of both LH and LS. The articulatory parameters studied were thus jaw opening and LA.

4.1.2. *Possible durational correlates*

The major durational correlates of focus identified in the study of reiterant speech were: focal lengthening, prefocal lengthening and what was called “lenghtening of initial lip closure” for the first [ma] in the focalized phrase. Similarly, in this corpus, focal and prefocal duration were measured, as well as the duration of the first phoneme of the focused sequence. This last parameter will thereafter be referred to as “first segment duration”. In so doing , we wanted to find out if the lengthening of the initial lip closure measured for reiterant speech was only an artifact of the syllable used or a general correlate of contrastive focus in French.

4.2. Results

4.2.1. *Articulatory correlates*

The maxima over each syllable were automatically detected for the two articulatory parameters, and the rise from broad focus to narrow focus was computed. Then the mean of these rises was computed over all the phrase (S, V or O).

Lip area (LA)

The mean increase of S from a broad focus to a focus condition is of 36.4%. This value can seem low but considering the diversity of the vowels of the corpus, it is interesting. Figure 3 shows the grand mean of the percentages of the rise of LA over each syntactic phrase and over all the identical syntactic phrases of the corpus. For example, the 1st column was computed in the following way: the peaks of lip area were detected for all the syllables of the subject of the broad focused utterances, the means over each subject were then calculated and the means of these means were plotted. ANOVA tests showed that for S, V and O the hypothesis of equality of the means of the four conditions can be rejected (S: F(2,42) = 12.21 p < 0.01, V: F(2,42) = 9.48 p < 0.01, O: F(2,42) = 50.19 p < 0.01). Student tests enable us to say that the grand means of the rises for S, V and O under focus is significantly (p < 0.01) greater than 0 (broad focus condition). There is thus a significant increase in lip area from broad focus to focus (see Figure 3).

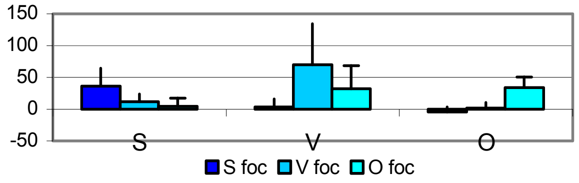

Figure 3: *Grand mean of the max of LA (cm²) over S, V & O.*

Jaw opening

The mean increase in jaw opening from a broad focus to a focus condition is of 51.72%. However, ANOVA tests showed that for S, V and O the hypothesis of equality of the means of the four conditions cannot be rejected. This implies that the mean increase in jaw opening due to focus is not significant throughout the corpus.

Post-focal hypoarticulation

The phrases after focus are hypoarticulated compared to the same phrases in the broad focused phrases: reduced lip area and jaw opening.

4.2.2. *Durational correlates*

Focal lengthening

As explained in the tonal validation of the corpus (2.3), the mean duration of the focused syllables was measured and compared to that of the same syllables in the unfocused versions of the utterances. The mean duration of the syllables was significantly higher (epsilon test, p < 0.01) for the focused condition. The mean lengthening from the broad focus case to the focus case is of 33.71%.

Prefocal lengthening

The duration of the syllable preceding the focused phrase was measured and compared to that of the same syllable for a broad-focused utterance. The duration of the last syllable of a phrase was significantly higher (epsilon test, p < 0.01) when the following phrase was focused. The mean lengthening is of 19.63%.

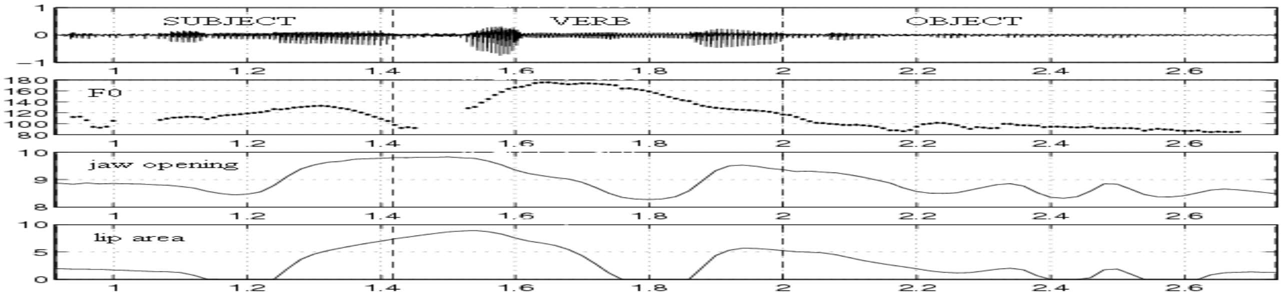

Figure 4: *Traces of the acoustic signal, the jaw opening (cm) and the lip area (cm²) as a function of time (s). The utterance pronounced was [Romain]$_S$ [RANIMA] $_V$ [la jolie maman] $_O$. The focal constituent is delimited by the dotted lines.*

First segment lengthening

As explained in 4.1.2., we measured the duration of the first phoneme of the focused phrase and compared it to that of the same first segment of the same phrase in the unfocused version of the utterance. It showed that the first segment was significantly lengthened (epsilon test, $p < 0.01$) when the phrase it belongs to was focused. The mean lengthening measured was of 59%. The first segment is therefore more lengthened than the rest of the focused phrase (only 33.71%).

4.2.3. *Sketch of a model of the visible correlates of contrastive focus in French*

As can be observed on the example presented in Figure 4, it seems that contrastive focus in French is characterized by a significant increase in lip area as well as in the durations of the prefocal syllable, of the first segment of the focused phrase and (although less so) of the rest of the focused syllables. The post focal phrase(s) are also hypo-articulated.

5. Conclusions

A preliminary study [17] on reiterant speech ([ma] repetitions) had shown that focus implied a) a greater jaw opening and velocity b) a longer initial lip closure c) a lengthening of the prefocal syllable and of the focal syllables and d) a hypo-articulation of the post-focal sequence. For real speech, it was found that the lip area was the most significant parameter. This is consistent with the findings for reiterant speech. For [ma]s, hyper-articulation was always achieved through a larger jaw opening, but for real speech and depending on the syllables, it could be achieved either through lip opening or lip spreading. The variations of lip area represent both those of lip opening and lip spreading. It was also found that the lengthening of the prefocal syllable and of the focal syllables was a significant visible correlates of contrastive focus as had been found for reiterant speech. We also found a post-focal hypo-articulation. The lengthening of the first segment of the focal phrase (compared to a broad focus condition) was found to be highly significant. This correlate is actually linked to the duration of the initial lip closure found for the reiterant speech. In [18], it was shown that the correlates found for reiterant speech were very well perceived (correct identification of focus for 86% of the cases for a chance level of 25%). It was also proved that the correlates perceived were probably those identified in the preliminary production study described above. The same tests were conducted for real speech and similar results were found (71.45% of correct answers). This test is described in another paper submitted to this same conference.

6. Acknowledgments

We thank G. Rolland for designing and recording the corpus and C. Savariaux and A. Arnal for their technical help. We thank P. Welby for her comments on our work.

7. References

[1] Loevenbruck H., 1999. An Investigation of Articulatory Correlates of the Accentual Phrase in French. *Proceedings of ICPhS'99*. San Francisco, 1, 667-670.

[2] Tabain M., 2003. Effects of prosodic boundary on /aC/ sequences: articulatory results. *Journal of the Acoustical Society of America*, 113(5), 2834-2849.

[3] Burnham D., 2001. Visual discrimination of Cantonese tones by tonal but non-Cantonese speakers and by non-tonal language speakers. *Proceedings of AVSP'01*. Aalborg, Denmark, 155-160.

[4] Granström B., House D. & Lundeberg M., 1999. Prosodic Cues in Multimodal Speech Perception. *Proceedings of ICSLP'99*. San Francisco, 1, 655-658.

[5] De Jong K., 1995. The supraglottal articulation of prominance in English: linguistic stress as localized hyper-articulation. *Journal of the Acoustical Society of America* 97, 491-504.

[6] Bernstein L.E., Eberhardt S.P. & Demorest M.E., 1989. Single-channel vibrotactile supplements to visual perception of intonation and stress. *Journal of the Acoustical Society of America* 85, 397-405.

[7] Keating P., Baroni M., Mattys S., Scarborough R., Alwan A., Auer E.T. & Bernstein L.E., 2003. Optical Phonetics and Visual Perception of Lexical and Phrasal Stress in English. *Proceedings of 15th ICPhS*. Barcelona, 2071-2074.

[8] Jun S.-A., Fougeron C., 2000. A Phonological Model of French Intonation. In *Intonation: Analysis, modelling and technology*, A. Botinis (Ed.). Dordrecht: KAP, 209-242.

[9] Jun S.-A., Fougeron C., 2002 ?. Realizations of Accentual Phrases in French Intonation. *Probus* 14, 147-172.

[10] Beckman M. E., Hirschberg J. & Shattuck-Hufnagel S.. The original ToBI system and the evolution of the ToBI framework. *Prosodic typology and transcription: a unified approach*. S.-A. Jun (ed.), Oxford University Press.

[11] Liberman M., Pierrehumbert J., 1984. Intonational invariance under changes in pitch range and length. In *Language sound to structure: studies in phonology presented to Morris Halle by his teacher and students.* Aronoff M. & Oehrle R. (eds.). MIT Press, 157-233.

[12] Di Cristo A. 1998. Intonation in French. In *Intonation systems: a survey of twenty languages*, Hirst D. & Di Cristo A. (Eds.). Cambridge University Press, 195-218.

[13] Touati P., 1987. Structures prosodiques du suédois et du français. *Working Paper 21*. Lund University Press.

[14] Clech-Darbon A., Rebuschi G. & Rialland A., 1999. Are there Cleft Sentences in French?. In *The Grammar of Focus*. Tuller L. & Rebuschi G. (Eds), Amsterdam: Benjamins, 83-118.

[15] Lallouache M.-T., 1991. *Un poste Visage-Parole couleur. Acquisition et traitement automatique des contours de lèvres*. PhD Thesis, Institut National Polytechnique de Grenoble.

[16] Audouy M., 2000. *Traitement d'images vidéo pour la capture des mouvements labiaux*. Final engineering report, Institut National Polytechnique de Grenoble.

[17] Dohen M., Loevenbruck H., Cathiard M.A., Schwartz J.L., 2003. Potential Audiovisual Correlates of Contrastive Focus in French. *Proceedings of Eurospeech'03*. Geneva, 145-148.

[18] Dohen M, Loevenbruck H., Cathiard M.A., Schwartz J.L., 2003. Audiovisual Perception of Contrastive Focus in French. *Proceedings of the AVSP'03 Conference*. St Jorioz, France, 245-250.