



HAL
open science

A Multi-Measurement Approach to the Identification of the Audiovisual Facial Correlates of Contrastive Focus in French

Marion Dohen, H el ene Loevenbruck, Hill Harold

► **To cite this version:**

Marion Dohen, H el ene Loevenbruck, Hill Harold. A Multi-Measurement Approach to the Identification of the Audiovisual Facial Correlates of Contrastive Focus in French. Auditory-Visual Speech Processing Workshop 2005, Jul 2005, Vancouver Island, Canada. p. 115-116. hal-00370952

HAL Id: hal-00370952

<https://hal.science/hal-00370952>

Submitted on 25 Mar 2009

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 destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv es.

A MULTI-MEASUREMENT APPROACH TO THE IDENTIFICATION OF THE AUDIOVISUAL FACIAL CORRELATES OF CONTRASTIVE FOCUS IN FRENCH

*Marion Dohen**, *Hélène Loevenbruck** and *Harold Hill*[°]

* Institut de la Communication Parlée, Grenoble, France
 ° Human Information Science Research Labs, ATR, Kyoto, Japan
 dohen@icp.inpg.fr

ABSTRACT

The aims of this study are twofold. We first seek to validate a previous study conducted on one speaker. It identified lower face visible articulatory correlates of contrastive focus in French for real (non-reiterant) speech. We thus conducted the same study for another speaker. Our second goal was to enlarge the set of cues measured to other facial movements. To do so we used a complementary measurement technique (Optotrak) on the same corpus. The articulatory measurements showed that there was a set of consistent visible articulatory correlates of contrastive focus in French across speakers: a) an increase in lip area and jaw opening on the focused item b) a lengthening of the focal syllables c) a post-focal hypo-articulation. However, there are also some speaker specific strategies. The first speaker showed an anticipation strategy but the second speaker did not. The Optotrak measurements confirmed that: a) focus corresponded to an increase in lip height and/or in lip width b) the post-focal sequence was hypo-articulated. We also found that: a) the focused item was often correlated with a slight head movement (nod) b) facial movements such as cheek movements were amplified when there was focus and c) eyebrow movements were not correlated with focus.

1. INTRODUCTION

Contrastive focus is used to emphasize a word or group of words in an utterance as opposed to another. This study deals with the description of audiovisual prosodic contrastive focus in French ("XXXf a mangé la pomme." XXXf ate the apple.). Only very few studies have dealt with the possible visible correlates of contrastive focus [e.g. 1] and none have been done for French.

We have suggested in previous studies ([2,3]) that there were lower face articulatory visual cues to

contrastive focus in French. We found that contrastive focus in one speaker of French was characterized by an increase of 36.4% in lip area. Moreover, when a phrase was focused, it was observed that the following phrases were hypo-articulated (reduced lip and jaw movement and velocity). Durational measurements were also made since duration can be used as a visual cue as well. These measurements showed that the focused syllables were significantly lengthened by 34%. The first phoneme of the focused phrase was even more significantly lengthened (59%). It was also observed that the last syllable of a phrase was significantly lengthened when the following phrase was focused (19.6%). This was related to an anticipation strategy. However [2,3] were conducted for only one speaker. The present study thus aims at reproducing the analysis for another speaker in order to identify consistent cues from one speaker to another as well as variable articulatory strategies.

A visual only perception experiment was also conducted [4] and showed that the visible correlates identified were valid cues for the perception of contrastive focus (71.5% of correct answers on average for a chance level of 25%). A detailed analysis showed that some stimuli were well perceived even though the visible correlates identified above were not present. This suggested that other more subtle facial correlates may intervene. Studies on other languages have shown that other facial movements such as eyebrow movements [5] or head movements [6] could intervene. This is why it seems necessary to enlarge the set of facial movements measured by the use of a complementary technique (Optotrak).

2. EXPERIMENTAL METHOD

An audiovisual corpus was recorded from a male native speaker of French. It consisted of 13 sentences with a subject-verb-object (SVO)

structure and CV syllables. Four conditions were studied: narrow focus on each phrase (S, V, O) and broad focus (neutral version of the utterance). In order to trigger focus the speaker had to perform a correction task by focusing the phrase which had been mispronounced in a prompt. An acoustic validation was conducted. The corpus was recorded using two measurement techniques: a lip contour tracking device [7] designed at the *Institut de la Communication Parlée* (ICP, France) and a 3D optical tracking system: Optotrak (data collected at the Human Information Science Research Labs at ATR, Japan) which captures the light emitted by infrared emitting diodes (IREDs).

3. LIP TRACKING DATA

The lip area and protrusion (50Hz) were extracted from the video [7] and durational measurements were also made. The maxima of lip area for each phrase were detected and a normalization was performed so that each broad focus case corresponded to a lip area maximum of 1. This was done in order to compare lip area over the whole corpus which corresponded to real (non-reiterant) speech and thus featured high variability in lip area values from one syllable to another. It was found that lip area was increased on the focused item in 60% of the cases by an average of 23%. It was also observed that the duration of the focused syllables was lengthened by an average of 22% in 83% of the cases. Lip area in the post-focal sequence was reduced by an average of 24% in 95% of the cases. However no anticipation strategy (increase in lip area and duration on the syllable preceding the focused item) was found for this speaker, contrary to what we found for the previous study. An interesting fact is that in focused sequences, when lip area was not increased, most of the time duration was and vice-versa. This implies that this speaker could use either both correlates or only one of them.

4. OPTOTRAK DATA

A total of 24 IREDs were glued to the speaker's face. These IREDs were used to track the motion of the jaw, lips, cheeks, eyebrows. An additional 4 IREDs were attached to a head rig and were used to correct for head motion. IRED positions were sampled at 60 Hz and low-pass filtered. The acoustic signal was simultaneously recorded and sampled at 22 KHz.

Articulatory correlates: The articulatory measurements showed that focus was accompanied either by an increase in lip height or in lip width (or

both). We also found a post-focal hypo-articulation. This is consistent with the increase in lip area measured with the lip tracking system.

Other facial movements: We found that cheek and chin movements were amplified when the item considered was focused. However we found that there was no strong correlation between focus and eyebrow movements for this speaker.

Head movements : It was observed that focus was accompanied by a slight down-up head nod in most of the cases.

5. CONCLUSIONS

This study confirms that there are visible articulatory correlates to contrastive focus in French. Some of the correlates identified for this new speaker are consistent with those found for our first speaker [3]: a) increase in lip area and duration on the focused item b) post-focal hypo-articulation. However, some strategies (e.g. pre-focal anticipation) differ. The Optotrak data enabled us to enlarge the number of possible visible correlates of contrastive focus in French. We observed that focus is correlated to: a) a slight head nod; b) cheek and chin movements of greater amplitude; c) but no eyebrow movements. The same recordings will be made for other speakers to confirm these findings.

6. REFERENCES

- [1]. Keating, P., Baroni, M., Mattys, S., Scarborough, R., Alwan, A., Auer, E.T. & Bernstein, L.E., Optical Phonetics and Visual Perception of Lexical and Phrasal Stress in English, *Proceedings of ICPhS 2003*, 2071-2074.
- [2]. Dohen M., Loevenbruck H., Cathiard M.-A. & Schwartz J.-L., Visual perception of contrastive focus in reiterant French speech, *Speech Communication* 44, 155-172, 2004.
- [3]. Dohen M., Loevenbruck H., Cathiard M.-A. & Schwartz J.-L., Identification of the possible visible correlates of contrastive focus in French, *Proceedings of Speech Prosody 2004*, Nara (Japan), March 23-26, 2004.
- [4]. Dohen M., Loevenbruck H., Cathiard M.-A. & Schwartz J.-L., Can we see focus? A visual perception study of contrastive focus in French, *Proceedings of Speech Prosody 2004*, Nara (Japan), March 23-26, 2004.
- [5]. Granström, B., House, D. & Lundeberg, M., Prosodic Cues in Multimodal Speech Perception, *Proceedings of ICSLP'99*, 1, 655-658, 1999.
- [6]. Munhall, K.G., Jones, J.-A., Callan, D., Kuratate, T. & Vatikiotis-Bateson, E., Visual prosody and speech intelligibility: Head movement improves auditory speech perception, *Psychological Science*, 15(2), 133-137, 2004.
- [7]. Audouy, M., *Traitement des images vidéo pour la capture des mouvements labiaux*, Final engineering report, Institut National Polytechnique de Grenoble, 2000.