

The influence of the communicative relationship between manual gesture and speech on their temporal coordination in a designation task

Benjamin Roustan and Marion Dohen

1 Introduction

When interacting with others, we often produce manual gestures at the same time as we speak and these manual gestures are coordinated to speech. The framework of deixis or designation is particularly interesting to study the coordination of speech and manual gestures. It is indeed possible to designate with the hand by finger pointing and via speech using demonstratives or focus. Finger pointing is used to draw (focus) attention towards an object in the concrete or the abstract space. Prosodic focus is used to put forward a word or a group of words as being the most informative of the utterance. Pointing and prosodic focus are thus linked and achieve the same communicative function.

Several studies examined temporal coordination in multimodal deixis expressed by finger pointing and accentuation (see *e.g.* de Ruiter, 1998; Rochet-Capellan et al., 2008). In a previous study (hereafter referred to as *exp1*, see Roustan & Dohen, 2010), we found that prosodic focus "attracts" manual gestures, and especially the pointing gesture which achieves the same communicative function as focus. The "most relevant" parts of the pointing gesture occur within the focused constituent.

An important observation which can be made from most of the studies on speech/gesture coordination is that the results are highly variable from one study to another. The nature of the communicative relationship between gesture and speech may be one factor explaining those discrepancies. The aim of this study is to further explore the possible modulation of speech / gesture coordination by the communicative relationship between gesture and speech. In order to do so, we designed a paradigm very close to that used in our *exp1* in which we varied the communicative relationship between the pointing gesture and speech within the designation framework. In *exp1*, the pointing gesture and prosodic focus designated exactly the same object while in the present study (hereafter referred to as *exp2*), prosodic focus designated only one aspect of the information carried by both the entire spoken utterance and the visual target pointed at.

2 Methodology

Ten right-handed native adult French speakers participated in the experiment.

2.1 Design & Setup – As in *exp1*, and in order to address complementary research questions, we studied different types of manual gestures (pointing *vs* beats *vs* button pressing) but for the sake of clarity and conciseness regarding the question addressed here, the presentation of the results will be limited to the pointing gesture. Two narrow prosodic focus conditions were explored: subject (SF) *vs* object (OF) focus. A correction task was used to elicit the production of prosodic focus. Participants sat in front of a screen on which visual targets appeared. They heard an audio prompt and saw two images on the screen which naturally elicited a correction of the audio prompt (see Table 1). Participants were asked to point at the corresponding visual target while they performed the correction. The corpus consisted of six simple subject-verb-object French utterances. We recorded participants' vocal productions (using a microphone) as well as the motion of their mouth and right hand (using the Optotrak motion capture system). Several events were labelled on the motion signals: apex (end of the extension movement for a pointing gesture) and beginning of return for manual gesture and articulatory lip vocalic targets (lip aperture/protrusion peaks). We also labelled acoustic cues: fundamental frequency and intensity peaks and syllable boundaries. All temporal data were normalized against the acoustic timing of the spoken utterance (beginning of the utterance = 0, end of the utterance = 1).

2.2 Difference with *exp1* – Table 1 shows the main differences between *exp1* and *exp2*. In *exp1*, the focused constituent exactly corresponded to the visual target pointed at (*e.g.* focus on "baby" in "Baba holds the baby" and pointing at an image of a baby). In *exp2*, the utterance carried the information of the visual target pointed at but only part of it was focused (*e.g.* focus on "balloon" in "The balloon is red" and pointing at an image of a red balloon). In *exp1*, the pointing gesture and prosodic focus designated the same thing but not in *exp2*. Moreover, in *exp2*, the object pointed at in both focus conditions (SF and OF) is the same. We therefore expected no temporal difference in the coordination between gesture and speech from SF and OF (unlike in *exp1*).

	Subject Focus		Object Focus	
	<i>exp1</i>	<i>exp2</i>	<i>exp1</i>	<i>exp2</i>
Audio Prompt	- Baba holds the baby. - She said: Mumu holds the baby ?	- The balloon is red.	- Baba holds the baby. - She said: Baba holds the present ?	- The balloon is green.
Images	Baba & Mumu 2 cartoon characters	red balloon & green candy	baby & present	red balloon & yellow candy
Participant	- BABA holds the baby.	- The BALLOON is red.	- Baba holds the BABY.	- The balloon is RED.
Pointed at targets	Baba ●	red balloon ●	baby ●	red balloon ●

Table 1: Illustration of the experimental paradigm used and comparison with *exp1* (all speech material was translated into English, small capitals stands for focused parts of speech)

3 Results

We studied the temporal coordination between pointing gesture and speech productions using paired *t*-tests, all the results are reported with the corrected number of degrees of freedom. The dependant variables used were apex and return times. The general *t*-test on the data reveals a significant main effect of focus condition, reflecting the fact that apex and return times are larger for OF than for SF. Therefore, it appears that focus attracts the pointing gesture, just as in *exp1*. However, further analysis of the data shows that the participants can be split into two groups. The first one (Group 1; 4 participants) displays the same behavior as in *exp1*: the *focus* condition affects the timeline of gesture production (apex: $t(6) = -2.7, p < .05$, means: $m_{FS} = 0.39, m_{FO} = 0.59$; return: $t(5.75) = -2.6, p < .05$, means: $m_{FS} = 0.65, m_{FO} = 1.02$). For the other one (Group 2; 6 participants), the *focus* condition does not affect the timeline of gesture production (apex: $t(9.12) = -.5, p = .6$, means: $m_{FS} = 0.40, m_{FO} = 0.44$; return: $t(8.92) = -.65, p = .53$, means: $m_{FS} = 0.61, m_{FO} = 0.70$).

The timing of the pointing gesture relative to speech was compared between *exp1* and *exp2* for both focus conditions. For both groups and in the SF condition, both the focused constituent and the pointing apex occurred later in *exp2* (systematic and equal time shift). This is due to the fact that in *exp2* there was a determiner before the focused noun of the subject, which was not the case in *exp1*. Speech/gesture coordination was therefore similar in *exp1* and *exp2* for SF.

However, the same does not hold for OF in Group 2. The pointing gesture occurs earlier than in *exp1* and this cannot be entirely explained by a shift of the object itself: the shift of the pointing apex is twice as big as the shift of the beginning of the object ($t(10.76) = 2.6; p < .05$, means: apex shift= 0.15, focus shift= 0.061). For OF (Group 2), the pointing gesture is therefore shifted toward the subject compared to *exp1*.

4 Conclusions and discussion

The purpose of this study was to analyze the modulation of speech/gesture coordinative strategies by the communicative link between manual gestures and speech in a designation task. Whereas in a previous experiment (*exp1*) participants designated the same object through focus and manual pointing, in this experiment, participants designated through focus only part of what was designated by manual pointing.

Results showed that attraction of the manual gesture by focus was weaker than in *exp1* and that participants could be split into two groups. Some tend to coordinate the manual pointing gesture with focus just as in *exp1* but for the others (majority of participants) the coordinative pattern is not significantly different from one focus condition to another (subject vs object). These results show that, in the framework of designation, speech/gesture coordinative patterns are modulated by the communicative relationship between speech and gesture. The parts of speech and gesture that show overlap (see Rochet-Capellan et al., 2008) especially if they actually show the same thing.

References

- de Ruiter, J. (1998). *Gesture and speech production*. Unpublished doctoral dissertation, Catholic University of Nijmegen, Netherlands.
- Rochet-Capellan, A., Laboissière, R., Galvan, A., & Schwartz, J.-L. (2008, December). The speech focus position effect on jaw-finger coordination in a pointing task. *Journal of Speech, Language, and Hearing Research*, 51, 1507–1521.
- Roustan, B., & Dohen, M. (2010, May). Co-production of contrastive prosodic focus and manual gestures: Temporal coordination and effects on the acoustic and articulatory correlates of focus. In *Speech prosody 2010 conference*. Chicago, USA.