

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
601  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

# The speech focus position effect on jaw-finger coordination in a pointing task

Amélie Rochet-Capellan <sup>a)</sup>, Rafael Laboissière <sup>b,c,d)</sup>, Arturo Galván <sup>d)</sup>, Jean-Luc Schwartz <sup>a)</sup>

<sup>a)</sup> GIPSA-Lab, Département Parole et Cognition (ICP) UMR CNRS 5216 - INPG – Université Stendhal, Grenoble, France

<sup>b)</sup> INSERM, U864, Espace et Action, Bron, France

<sup>c)</sup> Université Claude Bernard, Lyon 1, IFR19, IFR25, UMR-S864, Bron, France

<sup>d)</sup> Max Planck Institute for Human Cognitive and Brain Sciences, Munich, Germany

Suggested running head: Focus position and jaw-finger coordination in a pointing task

<sup>1</sup>-Corresponding author: amelie.rochet-capellan@gipsa-lab.inpg.fr, tel. (+33)(0)476574827, fax. (+33)(0)4.76.82.43.35, INP Grenoble, 46 avenue Félix Viallet, 38031 GRENOBLE CEDEX 1, France

## 25 **Abstract**

26 **Purpose:** This paper investigates jaw-finger coordination in a task involving pointing to a  
27 target while naming it with a 'CVCV (e.g. /'papa/) vs. CV'CV (e.g. /pa'pa/) word. According to  
28 our working hypothesis, the pointing apex (gesture extremum) would be synchronized with the  
29 apex of the jaw opening gesture corresponding to the stressed syllable.

30 **Method:** Jaw and finger motions were recorded using Optotrak. The effects of stress position  
31 on jaw-finger coordination were tested across different target positions (near vs. far) and  
32 different consonants in the target word (/t/ vs. /p/). Twenty native Portuguese Brazilian  
33 speakers participated in the experiment (all conditions).

34 **Results:** Jaw response starts earlier and finger-target alignment period is longer for CV'CV  
35 words than for 'CVCV ones. The apex of the jaw opening gesture for the stressed syllable  
36 appears synchronized to the onset of the finger-target alignment period (corresponding to the  
37 pointing apex) for 'CVCV words, and with the offset of that period for CV'CV ones.

38 **Conclusions:** For both stress conditions, the stressed syllable occurs within the finger-target  
39 alignment period due to tight finger-jaw coordination. This result is interpreted as an evidence  
40 for an anchoring of the speech deictic site (*part of speech that shows*) in the pointing gesture.

41

42

43 **Key-words: Deixis, Pointing, Jaw, finger-jaw coordination, speech-hand coordination,**

44 **Optotrak measurements, focus, stress**

## 45 Introduction

46 Hand and mouth often work together in human behaviors, mainly in alimentation and  
47 communication. This link has motivated a large body of research. For example, Iverson and  
48 Thelen (1999) showed that spontaneous co-occurrence of hand and mouth movements  
49 appears right after birth. Then, around 6 to 8 months, hand and mouth start to mutually entrain  
50 each other in rhythmic activities characterized by manual and oral babbling. Gestures and  
51 speech are then produced sequentially at around 9 to 14 months and eventually synchronized  
52 at the age of 16 to 18 months. Interplay of hand and mouth motor control is also observed in  
53 adults' behavior. For example, when speakers open their mouth while grasping an object such  
54 as a fruit, the apertures of both the grasp and the mouth are adapted to the size of the object  
55 (Gentilucci *et al.*, 2001). The *observation* of an action realized by one part of the body, e.g.  
56 bringing a fruit to the mouth, also affects the *production* of an action realized by the other,  
57 e.g. uttering a syllable (see Gentilucci *et al.*, 2004; Gentilucci, 2003). Hand and mouth are  
58 also coupled in adults' rhythmic activities. For example, Kelso *et al.* (1981) found a 1 to 1  
59 ratio between the frequency of the repetition of the word "stack" and the simultaneous  
60 repetition of a flexion-extension motion of the index finger. In addition, the co-occurrence of  
61 hand and mouth movements is clearly observable in face-to-face communication. The origin  
62 of this co-occurrence seems to be motor rather than purely perceptual, since gestures are  
63 produced even in situations in which the interlocutor can't perceive them, such as in phone  
64 calls (Iverson and Goldin-Meadow, 1998).

65  
66 According to McNeill (2000), a variety of gestures can occur in communication, ranging from  
67 *gesticulations*, which are global, non-conventionalized, and speech-dependent, to *signs* in  
68 signed languages, which are segmented, analytic, conventionalized, and performed without  
69 speech. This paper focuses on a particular type of gesture that can accompany speech in

1  
2  
3  
4 70 communication, namely *pointing* gestures. The global aim here is to link deixis, the  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
71 component of language that allows referring to objects, with the capacity of synchronizing  
72 gesture and voice to show objects. This synchronization may depend on the properties of the  
73 motor coordination between hand and mouth, arising from pre-linguistic links between the  
74 two motor systems.

## 76 Pointing gesture and language

77 Our interest in pointing gestures and more particularly in their coordination with speech  
78 originates mainly from five observations reported in the literature. The first observation is that  
79 pointing gestures are the principal medium of shared attention, a basic function required for  
80 language acquisition (Tomasello *et al.*, 2007). The second observation is that pointing  
81 gestures appear to be universal (Butterworth, 2003), despite variability in the form of the  
82 gesture across cultures (Haviland, 2000; Wilkins, 2003). The third observation is that pointing  
83 gestures are the first and the dominant communicative actions in infant communication. At 12  
84 months, pointing gestures constitute 60% of infants' manual communicative gestures and are  
85 often accompanied by vocalizations (Butterworth, 2003). The fourth observation is that  
86 pointing gestures are at the cutting edge of language development. Goldin-Meadow and  
87 Butcher (2003) showed that the age at which children associate a pointing gesture with a word  
88 having complementary meanings determines the age of two-words productions (see also  
89 Volterra *et al.*, 2005; Pizzuto *et al.*, 2005 for similar conclusions). The fifth and last  
90 observation is that pointing gestures have been put forward as the canonical form of language  
91 demonstrative words (Haviland, 2000; Diessel, 1999). Drawing evidence from developmental  
92 and comparative psychology, Diessel (2006) argues that demonstrative words such as "this"  
93 or "that" serve the basic communicative function of joint (or shared) attention rather than a  
94 specific grammatical function. He provides evidences for considering demonstrative words as

95 particular linguistic objects, defending their universal character and especially their specific  
96 and close link with pointing gestures.

97  
98 Altogether, this body of research on the relationships between pointing gestures and language  
99 in general, and between pointing gestures and deixis in particular, led Abry *et al.* (2004) to  
100 consider the connection between hand and voice in deixis as a crucial step in language  
101 emergence. They proposed to derive speech and language from the necessity to localize the  
102 objects we need to talk about, which requires the hand and mouth coordination. Hence, the  
103 understanding of speech-showing and hand-pointing coordination could be considered as a  
104 key step to understand the emergence of language deixis.

105  
106 In this framework, this paper investigates the effect of the position of the emphasized part of  
107 speech, namely speech focus (*the part of speech that shows*), on jaw-finger coordination in a  
108 task involving pointing at a target with the hand/finger while naming it.

### 109 110 Processes underlying speech-pointing synchronization

111 At least since McNeill's work (1981), it is well known that speech and hand gestures are  
112 coordinated in on-line face-to-face interactions. This phenomenon has motivated studies  
113 about the processes involved in speech-hand coordination, around, among others, the question  
114 of the interaction vs. modularity of the two systems. Most often, these studies used a dual-task  
115 paradigm: the participant provided both a verbal and a gestural response to a stimulus. The  
116 hand and speech dynamics in this dual task are compared to hand dynamics in a gesture-only  
117 task and speech dynamics in a speech-only task. For example, in Holender (1980), the task was  
118 to name a letter that appeared on a screen and press a key while in Castiello *et al.* (1991), it  
119 was to pronounce "tah" in response to a visual stimulus that indicated an object to grasp.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

120 More in line with our concerns, Levelt *et al.* (1985), and later Feyereisen (1997) used the  
121 double task paradigm in order to study pointing gestures. The dual task was to point at an  
122 object with the hand while verbally designating it using a “that *object*” or “this *object*”  
123 utterance (e.g. “this lamp”). According to Levelt *et al.* (1985), pointing gestures present a  
124 double interest for the study of speech and hand synchronization: first, they are strictly  
125 dependent on the message being expressed and, second, the moment at which they reach their  
126 target (now referred to as the pointing apex) can be easily detected.

127  
128 Among others, Levelt *et al.*'s results showed that for utterances like “this lamp”, the voice-onset  
129 tends to be synchronized with the pointing apex. Hence, putting the target further from the  
130 subject delays both the pointing apex and the voice-onset. The voice-onset also occurs later in  
131 the dual task (when it is accompanied by the pointing gesture) than in the speech-only task. On  
132 the other hand, the timing of the pointing apex is essentially the same in both gesture-only and  
133 dual tasks. The authors interpret these results as evidence for an adaptation of speech  
134 commands to brachiomanual commands rather than the reverse. A delayed verbal response in a  
135 dual task as regards a speech-only task is also put forward in Holender (1980), Castiello *et al.*  
136 (1991), and Feyereisen (1997).

137  
138 However, all these studies measured the verbal response delay using the acoustic signal only,  
139 without considering the speech articulators. As we will discuss in the next section, the processes  
140 of speech-hand coordination might be better described and understood through the dynamical  
141 interplay between the orofacial articulation and the hand-finger systems.

## 142 143 **Jaw-hand coordination rather than voice-hand coordination**

144 The motivation for investigating the articulatory motions in speech-hand coordination stems

1  
2  
3  
4 145 from two kinds of arguments. First, at a methodological level, speech is also a gestural system  
5  
6 146 much like pointing. Following Stetson (1951), a great number of studies have focused on the  
7  
8 147 articulators' motions, characterizing speech as the outcome of a motor system. As suspected by  
9  
10 148 Castiello *et al.* (1991) and Holender (1980), some motor events might happen before the voice  
11  
12 149 onset. Hence, it is legitimate to investigate when articulators start to move relative to the  
13  
14 150 pointing gesture. In addition, at a theoretical level, speech-pointing coordination has been  
15  
16 151 assumed to emerge in the course of ontogeny from a developmental meeting between the jaw  
17  
18 152 and arm-hand motor control (Ducey-Kaufmann *et al.*, in press). According to MacNeilage and  
19  
20 153 Davis's Frame-then-Content scenario of speech development, speech motor control begins in  
21  
22 154 young babies with the mastering of the opening/closing oscillations of the jaw, which provides  
23  
24 155 the *speech-frame* (MacNeilage and Davis, 2000). The independent and coordinated control of  
25  
26 156 the tongue and the lips (the *content*) would be mastered later. In this frame-then-content  
27  
28 157 sequence experimentally observed in the course of ontogeny (Munhall and Jones, 1998; Green  
29  
30 158 *et al.*, 2002), the jaw is considered as the carrier of speech gestures. Yet, MacNeilage and Davis  
31  
32 159 did not consider the role of manual gestures in speech acquisition. Different studies put forward  
33  
34 160 a link between the motor control development of brachiomanual and orofacial gestures.  
35  
36 161 Supporting evidences for this link comes from the relationship between the frequencies of hand  
37  
38 162 and jaw oscillations in babbling (Petitto *et al.*, 2001; Iverson and Thelen, 1999; Ducey-  
39  
40 163 Kaufmann, 2007), what Ducey-Kaufmann *et al.* (in press) referred to as the *sign-frame* and  
41  
42 164 the *speech-frame*, respectively. According to them, the relationship between the frequencies  
43  
44 165 of the two systems would evolve toward a developmental meeting point between the *speech*  
45  
46 166 *frame* and the *sign frame*. This meeting point is suggested to be the basis of speech-hand  
47  
48 167 coordination and the background for the production of the first words. These two sets of  
49  
50 168 methodological and theoretical arguments, lead us to propose a jaw-hand rather than a voice-  
51  
52 169 hand investigation framework for studying speech and manual pointing coordination  
53  
54  
55  
56  
57  
58  
59  
60

## 170 An attraction between the speech focus and the hand focus

171 The question of interest in the present study concerns the candidate *sites* for the speech and  
172 pointing gesture coordination: which part of the hand gesture is synchronized with which part of  
173 the speech utterance? According to McNeill (1992), in speech-hand coordination, the hand  
174 gesture stroke is executed in synchrony with the semantically co-expressive word. Moreover,  
175 verbal deixis can be prosodic as well as grammatical (e.g. Løevenbruck *et al.*, 2005). When  
176 considering the communicative aim of speech-hand association in deixis, it seems reasonable to  
177 assume that *the part of the discourse that shows* should occur synchronously with *the part of the*  
178 *gesture that shows*. Thus, synchronization of speech and hand pointing in face-to-face  
179 communication could result in an attraction between the speech focus (the indexical word  
180 and/or the stressed part of the utterance) and the pointing focus (the moment at which the  
181 arm-hand-finger system is aligned with the target). This hypothesis is compatible with Levelt  
182 *et al.*'s (1985) results, which show a tendency towards synchrony between voice-onset  
183 corresponding to the demonstrative word ("this" or "that") and the hand-pointing apex.  
184 Nevertheless, Levelt *et al.* did not vary the position of the speech deictic site, which was,  
185 systematically at the beginning of the utterance (e.g. "this lamp" vs, "that lamp"). In this paper,  
186 we propose to vary the position of speech focus in a simple way by varying the stressed syllable  
187 in CVCV utterances. Our aim is to study how this variation influences the jaw-hand  
188 coordination in a task consisting in pointing to a target while naming it with a CVCV word. Our  
189 main hypothesis is that the hand pointing apex should be synchronized with the extremum (or  
190 apex) of the jaw opening gesture corresponding to the stressed syllable, either the first syllable  
191 in 'CVCV utterances (e.g. /'papa/) or the second syllable in CV'CV utterances (e.g. /pa'pa/). This  
192 alignment could be reached either through adaptation of the jaw movement to a constant hand  
193 movement in both 'CVCV and CV'CV sequences, or through a mutual adaptation, involving a  
194 modification of both jaw and hand motions across word stress conditions.

195

## 196 **Methods**

### 197 **Participants and language**

198 Brazilian Portuguese was chosen because it is one of the languages in which it is possible to  
199 find pairs of words that differ only by stress position (e.g. 'CVCV vs. CV'CV). The subjects  
200 were twenty native Portuguese Brazilian speakers (4 men and 16 women) aged 18 to 37 years  
201 (mean: 28.3, standard deviation: 5.3). They were paid 8 euros per hour for their participation.  
202 The participants were all right-handed, had no reported history of speech or hearing pathology  
203 and were unaware of the purpose of the experiment.

204

### 205 **Experimental design**

206 The experiment involved a hand-pointing task associated to the utterance of a CVCV disyllable.  
207 The main factor was the stress position in the CVCV disyllable: stress on the first vs. the  
208 second syllable, e.g. /'papa/ vs. /pa'pa/. The consonant was either /p/ or /t/. The vowel /a/ was  
209 selected because it requires a large jaw opening gesture. Moreover, two spatial targets were  
210 used for the pointing gesture (near vs. far). The variation of both the consonant and the target  
211 position contributed to focus participants' attention on the task. Hence, the experimental design  
212 consisted of three within-subjects two-level crossed factors: stress position (first vs. second),  
213 consonant (/t/ vs. /p/) and target position (near vs. far).

214

----- Figure 1 -----

215

### 216 **Procedure**

217 The participants were seated at a table. The targets to point at and the item to pronounce were  
218 projected simultaneously on a white screen in front of them using a projector (Figure 1, top).

1  
2  
3  
4 219 A black square pasted on the midline of the table, close to the participant's sagittal plane,  
5  
6 220 indicated the finger resting position. The participants were informed that a word and a red  
7  
8 221 smiley sign (the target) would appear on the screen. The target appeared to the participant's  
9  
10 222 right (Figure 1, bottom) either near (10 cm from midline) or far (50 cm from midline). In  
11  
12 223 order to make the joint gesture/pronunciation task more natural, participants were instructed  
13  
14 224 to use the word displayed as the name of the person represented by the smiley target.  
15  
16 225 Participants were instructed to simultaneously point with the index finger at and name the  
17  
18 226 target as soon as the color of the smiley sign changed from red to green. Prior to the  
19  
20 227 experiment, participants were briefly trained to become familiar with the task: they were  
21  
22 228 asked to simultaneously point at and name objects in the room. They also practiced reading  
23  
24 229 CVCV sequences aloud in order to make sure that they understood the stress instruction  
25  
26 230 properly. The experiment was divided into four blocks. One block contained 4 practice trials  
27  
28 231 followed by 40 experimental trials, 5 for each combination of stress position, consonant, and  
29  
30 232 target position. The order of the trials was randomized for each block and each participant.  
31  
32 233 Blocks were separated by 30s rest periods. In order to reduce anticipatory responses to the go  
33  
34 234 signal (smiley target becoming green), the red smiley duration was varied from trial to trial  
35  
36 235 (2.5 s mean, 0.15 s standard deviation, normally distributed). The green smiley target lasted  
37  
38 236 on the screen for 1 s in each trial.  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51

## 52 238 Data recording and post-processing

53 239 Finger and jaw movements were recorded using Optotrak (Northern Digital, Inc.), an  
54  
55 240 optoelectronic position measurement system that tracks the three-dimensional motion of  
56  
57 241 infrared-emitting diodes (IREDs). The positions were sampled at 100 Hz. IRED locations are  
58  
59 242 illustrated on Figure 1 (top). Two IREDs were pasted on the tip of participants' right  
60  
243 forefinger, one on the middle of the nail and the other on the medial side next to the nail. In so

1  
2  
3 244 doing, at least one of the IREDs was visible by the cameras during the pointing movement,  
4  
5 245 even when participants supinated their hands at the motion apex. A third IRED was attached  
6  
7  
8 246 to the participants' chin. It tracked a flesh point rather than the jaw itself. However,  
9  
10 247 considering the phonetic material in question (stop consonants associated with an open  
11  
12 248 vowel), the motion of this flesh point is a relevant indicator of jaw motion. Head motion was  
13  
14 249 measured by three IREDs attached to a plastic triangle, which was fixed by a strap around the  
15  
16  
17 250 participant's head. The coordinates of the moving IREDs were projected into a fixed  
18  
19 251 referential, defined by three IREDs pasted on the table. Jaw position was then computed in  
20  
21 252 the head moving reference frame. Principal Component Analysis (PCA) was applied separately  
22  
23 253 to each of the three 3-D trajectories of the two finger IREDs and the jaw IRED. The first  
24  
25 254 principal component explained most of the variance for each IRED and for all participants:  
26  
27 255 98.8% (standard error = 0.2%) and 98.3% (standard error = 0.3%) for the two finger IREDs and  
28  
29 256 95.6% (standard error = 0.5%) for the jaw IRED. This component was chosen to represent  
30  
31 257 finger and jaw movements. Signals were low-pass filtered at 15 Hz with a Butterworth. The  
32  
33 258 sound was simultaneously recorded and sampled at 16 kHz.

34  
35 259  
36  
37 260 The recorded utterances were checked against the correct phonetic and stress instructions.  
38  
39 261 Trials with speech production errors were excluded from the dataset (in average, 4.1 trials per  
40  
41 262 subject, with a maximum of 11 errors for one participant, see table 1, first row). Correct trials  
42  
43 263 for which jaw or finger gestures were initiated before the go-signal were also discarded. This  
44  
45 264 mainly concerned one participant who made 39 initiation errors (Table 1, second row). For  
46  
47 265 the remaining data, trials for which the two finger IREDs were partially hidden from the  
48  
49 266 Optotrack cameras were not considered for the analysis (8.2 trials in average, see table 1, third  
50  
51 267 row). By default, the middle IRED was chosen for the analysis. When this IRED was masked  
52  
53 268 and the left one was visible, the left IRED was taken for the analysis (the mean correlation

269 between the two finger IREDs is above .99 for each of the three coordinates x, y, and z).  
270 Similarly, trials for which the jaw IRED was partially hidden were discarded from the  
271 analysis, which mainly concerned one participant for 12 trials (Table 1, fourth row). A three  
272 within-subject factors ANOVA shows that stress position, consonant and target position do  
273 not significantly affect either the number of utterance and initiation errors, or the number of  
274 trials with hidden finger or jaw IREDs.

275

## 276 Labeling and Measurements

277 Figure 2 displays an example of signals and labeling for a /pa'pa/ trial. Onset and offset events  
278 of the finger and jaw movements correspond to 10% of peak velocity. All times correspond to  
279 the elapsed times from the go signal to the event.

280 ----- Figure 2 -----

281

282 For the jaw (Figure 2, second row), the analysis was focused on the two opening strokes for  
283 the /a/ vowels.  $J_{I1}$  and  $J_{A1}$  are the initiation and apex times for the first opening stroke (stroke  
284 onset and offset, respectively). Similarly,  $J_{I2}$  and  $J_{A2}$  are the initiation and apex times for the  
285 second opening stroke. For the finger, the trajectory can be split into three parts: the forward  
286 stroke, the plateau and the return stroke (see Figure 2, last row). The forward stroke  
287 corresponds to the pointing gesture toward the target. It starts at  $P_I$  (initiation time of pointing  
288 gesture) and ends at  $P_A$  (apex time of pointing gesture). The pointing plateau is the amount of  
289 time during which the finger remains pointed at the target. It starts at  $P_A$  and ends at  $P_R$  (onset  
290 time of the return stroke). The return stroke corresponds to the finger return back movement  
291 to its rest position. This stroke was not considered here since it is not really part of the  
292 pointing task. Note that Levelt *et al.* (1985) only investigated the pointing forward stroke.  
293 Yet, the observation of a plateau in the apex position shows that the finger-pointing task does

294 not end with the alignment of the finger and the target but rather with the onset of the return  
295 stroke. In our study, for the finger motion analysis we considered both the forward stroke and  
296 the plateau.

297

## 298 Hypotheses

299 Under the assumption that there is a tight temporal coordination of jaw and finger pointing  
300 gestures, the apex of the jaw opening gesture corresponding to the stressed syllable should be  
301 synchronized with the pointing apex. This might induce a significant effect of the stress  
302 position on the delay between the apex of the first jaw opening gesture,  $J_{A1}$ , and the apex of  
303 the pointing gesture,  $P_A$ , that should be shorter in the first-syllable stress condition than in the  
304 second-syllable stress one. The increase in the delay between  $J_{A1}$  and  $P_A$  in the second-  
305 syllable stress condition would result in the apex of the second jaw opening gesture,  $J_{A2}$ ,  
306 being closer to  $P_A$ . This should occur regardless of the target position and of the utterance  
307 consonant. Hence, the effect of stress position on the timing of the pointing apex relatively to  
308 the two jaw opening gestures apices should be similar for the two target positions and the two  
309 consonants. Finally, the effect of stress position on jaw-finger coordination could result from  
310 either an adaptation of jaw movement to hand movement (or the converse) or a mutual  
311 adaptation between the two motor systems. This can be evaluated through the analysis of the  
312 effect of stress position on timing, duration and amplitude of finger and jaw gestures.

313

## 314 Results

315 Figure 3 shows the mean temporal position of jaw and finger events, computed for the  
316 20 participants, depending on stress position, consonant and target position. The effects of the  
317 experimental variations on each measured variable were tested using three-way (stress

318 position, consonant, and target position) within-subject ANOVAs. The p-value for the  
319 significance level was fixed at 0.05.

320 ----- Figure 3 -----

321

## 322 Absolute and relative timing of finger and jaw motions

### 323 *Initiation times*

324 The pointing motion starts on average 322 ms after the go signal (see  $P_I$  values on Figure 3).

325 There is no significant main effect of any of the three factors on  $P_I$ . Contrastively, the

326 initiation of the first jaw stroke ( $J_{II}$ ) occurs significantly earlier in the second-syllable stress

327 condition (476 ms) than in the first-syllable stress one (539 ms) ( $F(1, 19) = 15.9, p < .001$ ),

328 and for /t/ (488 ms) as compared to /p/ (527 ms), ( $F(1, 19) = 13.5, p < .01$ ). Neither the

329 interaction between stress position and consonant, nor that between stress position and target

330 position is significant. Hence, the onset of the jaw movement ( $J_{II}$ ) depends on stress position

331 and occurs later than the onset of the finger movement ( $P_I$ ). Analysis of  $J_{II} - P_I$  delay shows

332 that  $J_{II}$  is closer to  $P_I$  in the second-syllable stress condition (149 ms) than in the first-syllable

333 stress one (223 ms) ( $F(1, 19) = 14.1, p < .01$ ), and for the /t/ (166 ms) as compared to /p/

334 (206 ms) ( $F(1, 19) = 13.3, p < .01$ ). Furthermore,  $J_{II} - P_I$  is significantly shorter for the near

335 (176 ms) than for the far (195 ms) target position ( $F(1, 19) = 10.9, p < .01$ ). Interaction effects

336 on  $J_{II} - P_I$  are not significant. Hence, the jaw starts to move after the finger but significantly

337 closer to the beginning of the finger movement when stress is on the second rather than on the

338 first syllable. Moreover, the effect of stress position on the delay between  $P_I$  and  $J_{II}$  results

339 from an effect on  $J_{II}$ : the stress position has no significant effect on  $P_I$ .

### 340 *Apices times*

341 The pointing apex ( $P_A$ ) occurs 16 ms later in the far (749 ms) than in the near (733 ms) target

342 condition. The effect of target position is significant ( $F(1, 19) = 9.2, p < .01$ ). Stress position

343 and consonant type do not have any significant effect on  $P_A$ . By contrast, the stress position  
 344 significantly influences the apices times of the jaw opening gestures. The apex of the first jaw  
 345 opening gesture ( $J_{A1}$ ) occurs 101 ms earlier in the second-syllable stress condition (620 ms)  
 346 than in the first-syllable stress one (721 ms), ( $F(1, 19) = 54.0, p < .0001$ ). A similar delay  
 347 across stress conditions is observed for the apex of the second jaw opening gesture ( $J_{A2}$ )  
 348 which occurs on average 108 ms earlier in the second-syllable stress condition (900 ms) than  
 349 in the first-syllable stress one (1008 ms), ( $F(1, 19) = 54.8, p < .0001$ ). Target position effect is  
 350 also significant for both  $J_{A1}$  ( $F(1, 19) = 7.1, p < .05$ ) and  $J_{A2}$  ( $F(1, 19) = 5.9, p < .05$ ). The  
 351 apices of both jaw opening gestures occur earlier in the near (665 ms for  $J_{A1}$  and 949 ms for  
 352  $J_{A2}$ ) than in the far (677 ms for  $J_{A1}$  and 959 ms for  $J_{A2}$ ) target condition. The increase of the  
 353 target distance thus delays the apices of the two jaw opening gestures by about 10 ms. This  
 354 delay is equivalent to the non-significant delay observed for the onset of the jaw movement,  
 355  $J_{I1}$ , and seems to partly compensate for the 16 ms delay observed for the pointing apex,  $P_A$ .  
 356 Finally, the consonant has no significant effect on the apices times of the two jaw opening  
 357 gestures.

358  
 359 In addition, the delay observed for  $J_{A1}$  in the first-syllable stress condition compared with the  
 360 second-syllable one results in  $J_{A1}$  being closer to  $P_A$  in the first case. The study of  $J_{A1} - P_A$   
 361 shows that  $P_A$  occurs about 11 ms after  $J_{A1}$  in the first-syllable stress condition and about  
 362 129 ms after  $J_{A1}$  in the second-syllable stress one. This effect of stress position is significant  
 363 ( $F(1, 19) = 52.3, p < .0001$ ). On the other hand, neither the target position nor the consonant have a  
 364 significant effect on  $J_{A1} - P_A$ . In order to better characterize the temporal relationships between  
 365 finger and jaw apices, we computed the position of pointing apex,  $P_A$ , relatively to the apices  
 366 of the two jaw opening gestures,  $J_{A1}$  and  $J_{A2}$ :

367 
$$P_{A/J} = \frac{P_A - J_{A1}}{J_{A2} - J_{A1}}, \text{ in percent.}$$

368 A value of 0% for  $P_{A/J}$  corresponds to a case for which  $P_A$  occurs at the same time as  $J_{A1}$ . A  
369 value of 100% corresponds to a case for which  $P_A$  is synchronized with  $J_{A2}$ . Figure 4 (left)  
370 shows  $P_{A/J}$  means and standard errors against stress conditions. It appears that  $P_{A/J}$  mean is  
371 greater when the second syllable is stressed (47%) rather than the first one (5%)  
372 ( $F(1, 19) = 66.0, p < .0001$ ). The effects of target position and consonant type as well as the  
373 interactions are not significant. In line with the jaw-finger synchronization hypothesis, stress  
374 position influences the relative times of the apices of the jaw opening gestures and of the  
375 pointing gesture. While the pointing apex is very close in time to the apex of the first jaw  
376 opening gesture when the first syllable is stressed, it occurs at about an equal distance in time  
377 from the apices of the two jaw opening gestures when the second syllable is stressed.  
378 Strikingly, in this last condition, the pointing apex seems closely synchronized with the  
379 initiation of the second opening gesture of the jaw,  $J_{I2}$ . In average,  $P_A$  occurs just 1 ms after  $J_{I2}$   
380 when the second syllable is stressed and 175 ms before  $J_{I2}$  when the first syllable is stressed  
381 (see Figure 3). Moreover, in the second-syllable stress condition, the apex of the second jaw  
382 opening gesture is very close to  $P_R$ , the onset of the pointing gesture return stroke (that is, the  
383 offset of the pointing plateau).

384 ----- Figure 4 -----

385 ***Offset of the pointing plateau ( $P_R$ ) relative to the jaw apices***

386  $P_R$  occurs later when the second syllable is stressed (906 ms) rather than the first one (860 ms)  
387 ( $F(1, 19) = 16.0, p < .001$ ) as well as when the target is far (901 ms) as compared to when it is  
388 near (864 ms) ( $F(1, 19) = 28.0, p < .0001$ ). The consonant factor as well as all the interactions  
389 is not significant. The pointing plateau duration ( $P_R - P_A$ ) is significantly greater when the  
390 second syllable is stressed (157 ms) rather than the first one (127 ms) ( $F(1, 19) = 14.1,$   
391  $p < .01$ ). It is also significantly greater when the target is far (152 ms) rather than near  
392 (132 ms) ( $F(1, 19) = 13.9, p < .01$ ). As for  $P_R$ , neither the consonant factor, nor the

393 interactions are significant. Analysis of the  $J_{A2} - P_R$  interval shows that  $P_R$  occurs 5 ms after  
 394  $J_{A2}$  in the second-syllable stress condition and 148 ms before  $J_{A2}$  in the first-syllable stress  
 395 condition. This effect of stress position on  $J_{A2} - P_R$  is significant ( $F(1, 19) = 89.3, p < .0001$ ).  
 396 The delay between  $J_{A2}$  and  $P_R$  is also greater when the target is near (85 ms) rather than far  
 397 (58 ms) ( $F(1, 19) = 16.9, p < .001$ ) but does not depend on the consonant. Interactions on  
 398  $J_{A2} - P_R$  are not significant. Similarly to what was done for the pointing apex, we computed  
 399 the position of  $P_R$  relative to the apices of the two jaw opening gestures:

$$400 \quad P_{R/J} = \frac{P_R - J_{A1}}{J_{A2} - J_{A1}}, \text{ in percent.}$$

401 Figure 4 (right) shows that  $P_{R/J}$  is close to 100% for the second-syllable stress condition  
 402 (104%) while it is close to 50% for the first-syllable stress condition (51%).  $P_{R/J}$  is also  
 403 greater when the target is far (82%) compared to when it is near (73%) and for /t/ (79%) as  
 404 compared to /p/ (76%). The three main effects are significant (stress position:  
 405  $F(1, 19) = 102.9, p < .0001$ ; target position:  $F(1, 19) = 12.4, p < .01$ ; consonant:  
 406  $F(1, 19) = 4.9, p < .05$ ) but none of the interactions are. When the first syllable is stressed, the  
 407 finger leaves its apex position before the apex of the second jaw opening gesture, at about an  
 408 equal distance in time from the apices of the two jaw opening gestures. Actually, as  
 409 mentioned in the previous section, in the first-syllable stress condition, the finger even leaves  
 410 its apex position 175 ms before the apex of the second jaw opening stroke. By contrast, in the  
 411 second-syllable stress condition, the finger leaves its apex position just after the apex of the  
 412 second jaw opening gesture is reached.

413  
 414 In summary, the pointing apex,  $P_A$ , is close to the apex of the first jaw opening gesture,  $J_{A1}$ ,  
 415 when the first syllable is stressed while it occurs at about the midpoint between the apices of  
 416 the two jaw opening gestures,  $J_{A1}$  and  $J_{A2}$ , when the second syllable is stressed. In the second-  
 417 syllable stress condition, the jaw events ( $J_{I1}, J_{A1}, J_{I2}$  and  $J_{A2}$ ) occur earlier than in the first-

1  
2  
3 418 syllable stress condition and the onset of the pointing gesture return ( $P_R$ ) is delayed. These  
4  
5 419 two facts result in  $J_{A2}$  being close to  $P_R$ . By contrast, the timing of the pointing forward stroke  
6  
7  
8 420 does not depend on stress position. The increase of the target distance delays the pointing  
9  
10 421 apex but does not significantly influence its time position relatively to the apices of the jaw  
11  
12 422 opening motions:  $P_{AJ}$  is approximately the same in the two target conditions. This originates  
13  
14 423 from the fact that the apices of the jaw opening gestures are delayed when the target is far  
15  
16 424 compared to when it is near. Regardless of the experimental conditions, the stress in speech  
17  
18 425 always occurs sometime during the pointing plateau, either at the plateau onset for 'CVCV  
19  
20 426 words or at the plateau offset for CV'CV words. This results in the entire pointing plateau  
21  
22 427 occurring before the onset of the second jaw opening stroke when the first syllable is stressed  
23  
24 428 and after it when the second syllable is stressed. More precisely, the pointing plateau occurs  
25  
26 429 during the jaw closure stroke after the stressed vowel in the first-stress syllable condition and  
27  
28 430 during the jaw opening stroke toward the stressed vowel when the second syllable is stressed.  
29  
30 431 The finger and jaw motion will now be investigated in detail in order to further characterize  
31  
32 432 the impacts of stress, target and consonant on each system separately.  
33  
34 433

## 434 Detailed description of finger forward and jaw opening strokes

### 435 *Finger forward stroke*

436 The previous analysis showed that the pointing onset,  $P_I$ , does not depend on the experimental  
437 condition while the pointing apex,  $P_A$ , occurs significantly later when the target is far than  
438 when it is near. This timing pattern implies a greater duration of the forward stroke (computed  
439 as  $P_A - P_I$ ) when the target is far (431 ms) rather than near (406 ms) ( $F(1, 19) = 26.3$ ,  
440  $p < .0001$ , see Figure 5, first row, right). Neither the effect of stress position, nor the effect of  
441 consonant on the duration of the pointing forward stroke are significant. The amplitude of this  
442 stroke (computed as the distance between the pointing spatial positions at  $P_A$  and  $P_I$ , Figure 5,

443 first row, left) is also greater when the target is far (423 mm) compared to when it is near  
444 (275 mm) ( $F(1, 19) = 2222.0, p < .0001$ ). The 2 mm difference observed in the first-syllable  
445 stress condition (348 mm) compared to the second-syllable stress one (350 mm) is also  
446 significant ( $F(1, 19) = 8.3, p < .01$ ) but does not significantly interact with target position. The  
447 consonant does not have a significant effect on the amplitude of the pointing forward stroke.  
448 Hence, the pointing forward stroke is mainly influenced by target position: the alignment of  
449 the finger with the target requires greater amplitude and duration when the target is far  
450 compared to when it is near. This leads to consider that the pointing forward stroke is “target-  
451 driven”: its main objective is the alignment of the finger with the target. Consequently, the  
452 speech system may have to adapt in order to achieve the relative timing pattern between the  
453 apices of finger and jaw opening gestures, as previously observed.

454 ---- Figure 5 ----

#### 455 *Jaw opening strokes*

456 As described above, the initiations and apices events of the jaw opening strokes clearly  
457 depend on the experimental conditions. The duration of the first opening stroke (computed as  
458  $J_{A1} - J_{I1}$ , Figure 5, second row, right) is 39 ms greater when the first syllable is stressed  
459 (183 ms) rather than the second one (144 ms) ( $F(1, 19) = 26.2, p < .0001$ ). It is also 37 ms  
460 greater for /t/ (182 ms) compared to /p/ (145 ms) ( $F(1, 19) = 14.5, p < .01$ ). The interaction  
461 between stress position and consonant is also significant: the effect of stress position is larger  
462 for /p/ (first-syllable: 168 ms, second-syllable: 122 ms) than for /t/ (first-syllable: 197 ms,  
463 second-syllable: 167 ms), ( $F(1, 19) = 5.5, p < .05$ ). However, target position does not  
464 significantly influence the duration of the first opening stroke. Similarly, the amplitude of the  
465 first jaw stroke (distance between the jaw spatial positions at  $J_{I1}$  and  $J_{A1}$ , Figure 5, second  
466 row, left) is greater when the first syllable is stressed (12.5 mm) than when the second one is  
467 (9.6 mm) ( $F(1, 19) = 21.1, p < .001$ ). Consonant also has a significant effect: the mean

1  
2  
3  
4 468 amplitude of the first opening gesture is greater for /p/ (11.7 mm) than for /t/ (10.3 mm)  
5  
6 469 (F(1, 19) = 9.7,  $p < .01$ ). Neither target position, nor the interactions is significant. Contrary to  
7  
8 470 the first opening stroke, the duration of the second opening stroke ( $J_{A2} - J_{I2}$ , Figure 5, third  
9  
10 471 row, right) is greater when the second syllable is stressed (152 ms) rather than the first one  
11  
12 472 (100 ms) (F(1, 19) = 42.1,  $p < .0001$ ) and for /p/ (138 ms) as compared to /t/ (114 ms)  
13  
14 473 (F(1, 19) = 47.1,  $p < .0001$ ). There is a significant interaction between stress position and  
15  
16 474 consonant (F(1, 19) = 8.2,  $p = .05$ ): the effect of stress position is greater for /t/ (first-syllable:  
17  
18 475 84 ms, second-syllable: 145 ms) than for /p/ (first-syllable: 117 ms, second-syllable: 159 ms).  
19  
20 476 Target position does not significantly influence  $J_{A2} - J_{I2}$ . A similar pattern is observed for the  
21  
22 477 amplitude of the second stroke (difference between jaw spatial positions at  $J_{I2}$  and  $J_{A2}$ ,  
23  
24 478 Figure 5, third row, left) which is greater when the second syllable is stressed (7.5 mm) rather  
25  
26 479 than the first one (3.2 mm) (F(1, 19) = 40.3,  $p < .0001$ ) and for /p/ (6.7 mm) compared to /t/  
27  
28 480 (4.0 mm) (F(1, 19) = 45.8,  $p < .0001$ ). There is also a significant interaction between stress  
29  
30 481 position and consonant (F(1, 19) = 5.7,  $p < .05$ ): the effect of stress position is greater for /t/  
31  
32 482 (first: 1.6 mm, second: 6.4 mm) than for /p/ (first: 4.8 mm, second: 8.7 mm). There is no  
33  
34 483 significant effect of target position. These results show that in order to produce stress as  
35  
36 484 instructed, speakers increase both the duration and the amplitude of the jaw opening stroke  
37  
38 485 corresponding to the stressed syllable. The specific articulatory configurations of /p/ and /t/  
39  
40 486 result in significant effects of the consonant factor on the amplitude and the duration of the  
41  
42 487 jaw opening strokes. However, these effects of consonant appear to have no influence on the  
43  
44 488 timing of the apices of the jaw opening gesture. Finally, the target position does not have a  
45  
46 489 significant effect on the amplitude and the duration of the jaw opening strokes: increase in  
47  
48 490 distance to the target induces a delay around 10 ms for all jaw events studied here. These  
49  
50 491 results support the view that the timing of the pointing apex relatively to the timing of the  
51  
52 492 apices of the jaw opening gestures mainly originates from an adaptation of the jaw.  
53  
54  
55  
56  
57  
58  
59  
60

493

494 In summary, the analyses show that jaw-finger coordination basically consists in  
495 synchronizing the pointing apex with the apex of the first jaw opening gesture when the first  
496 syllable is stressed, and synchronizing the onset of the finger return gesture with the apex of  
497 the second jaw opening gesture when the second syllable is stressed. This pattern seems to  
498 rely on four types of adaptation:

499 (1) An adaptation of the finger pointing forward stroke to the spatial target: amplitudes and  
500 durations increase when the target is far rather than near;

501 (2) An adaptation of the jaw to the phonetic goal: specific durations and amplitudes  
502 correspond to /p/ vs. /t/, and amplitudes and durations are greater for stressed vs. unstressed  
503 syllables;

504 (3) An adaptation of the jaw to the pointing forward stroke. When the first syllable is stressed,  
505 whichever the spatial target is, the apex of the jaw opening gesture corresponding to the  
506 stressed syllable is synchronized with the pointing apex. When the second syllable is stressed,  
507 the delay between the apices of the two jaw opening gestures decreases, possibly allowing  
508 synchronization between the pointing apex and the onset of the jaw opening gesture  
509 corresponding to the second syllable;

510 (4) An adaptation of the duration of the pointing plateau, possibly compensating for the  
511 incomplete finger-jaw apex synchrony when the second syllable is stressed. In this condition,  
512 the increase of the duration of the pointing plateau allows for the apex of the jaw opening  
513 stroke corresponding to the stressed syllable to occur within the finger to target alignment  
514 period.

515

516 These results will now be discussed in the light of previous observations on speech-hand  
517 coordination and within the framework of the jaw-hand/finger coordination hypothesis.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 518 519 **Discussion**

520 The change in stress position from a 'CVCV to a CV'CV clearly influences the jaw-finger  
521 coordination. However it does so though in a different way than we have predicted. There are  
522 bidirectional adaptations from the jaw to the hand and from the hand to the jaw, which are  
523 superimposed on the intrinsic target-driven behaviors corresponding to each system. We shall  
524 discuss these three components (intrinsic behavior, jaw to hand and hand to jaw adaptations)  
525 separately, and conclude on the way the two systems could be coupled in a dynamic system  
526 approach.

### 527 528 A pointing task and a speech task

529 Obviously, the first point that must be made in this discussion is that, apart from any link  
530 between speech and hand gestures, the hand and speech systems reach their respective targets  
531 regardless of the individual requirements on each system. In the pointing task, the target must  
532 be designated and the movement must be adapted to the target position, which involves larger  
533 and longer movements when the target is far compared to when it is near. In the speech task, a  
534 speaker has to utter the right consonants with the right stress pattern, which involves an  
535 increase of the amplitude and/or the duration of the opening gesture for the stressed syllable.  
536 The phonetic identity of the consonant (labial *vs.* coronal) influences the initiation times of  
537 the jaw opening strokes but not their apices' times. It also affects the amplitude and duration  
538 of the jaw opening strokes. Asymmetry in the jaw cycles for labial and coronal constrictions  
539 in CV syllables is a phenomenon also observed in the repetition of CVCV words when speech  
540 rate is increased. One possible explanation for this finding is that, for anatomical reasons, the  
541 position of the jaw has a stronger influence on the position of the lower lip than on the  
542 position of the tongue (Rochet-Capellan and Schwartz, 2007a).

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60543  
544 **The jaw meets the hand**

545 As regards the first and major connection between the two systems, it appears that speech  
546 adapts to the hand rather than the reverse and that this adaptation is related to the jaw  
547 movements. Our prediction was that the synchronization between the apex of the jaw opening  
548 stroke corresponding to the stressed syllable and the apex of the pointing gesture would be a  
549 stable attraction point for jaw-finger coordination. This appears to be the case only when the  
550 first syllable is stressed. In this case, the observed large delay between the beginning of the  
551 pointing gesture and the beginning of the first jaw opening gesture is in accordance with  
552 previous results obtained using the dual-task paradigm (see Introduction section). When the  
553 second syllable is stressed, the delay between the onsets of the pointing gesture and the jaw  
554 movement is reduced due to an early onset of the jaw movement. However, this is not enough  
555 for the apex of the pointing gesture to be synchronized with the apex of the second jaw  
556 opening gesture. It could seem surprising that when the second syllable is stressed, the onset  
557 of the jaw movement occurs 148 ms (on average) after the onset of the hand movement, while  
558 the apex of the second jaw opening stroke occurs 151 ms after the apex of the pointing  
559 gesture. Initiating the first jaw opening gesture at the same time as the pointing gesture could  
560 be sufficient for the apex of the second jaw opening gesture to be synchronized with the apex  
561 of the pointing gesture. The hand then compensates this failed synchronization during its  
562 return to rest phase, as we shall discuss later. However, in terms of jaw adaptation, the  
563 synchronization seems partly unachieved. This does not mean that there is no meeting point  
564 between the jaw and the hand when the second syllable is stressed. In this condition, the jaw  
565 produces the first cycle early enough to entirely incorporate the pointing plateau inside the  
566 second cycle. Therefore, the general working hypothesis that “*the part of the discourse that*  
567 *shows is related to the part of the gesture that shows*” is verified. The movement of the jaw is

1  
2  
3 568 organized such that the pointing plateau unambiguously occurs within the appropriate temporal  
4  
5 569 domain of the spoken utterance that is the jaw cycle corresponding to the stressed syllable. This  
6  
7  
8 570 specifies the “deictic sites” we were looking for.  
9

10  
11 571

12 572 There seems to exist another meeting point that we had not predicted: when the second syllable  
13  
14 573 is stressed, the finger reaches its apex more or less exactly at the same time as the jaw reaches  
15  
16  
17 574 its highest position, which is likely to closely correspond to the consonantal contact time for the  
18  
19 575 second syllable (between the lips for /p/, or between the tongue and the palate for /t/).  
20  
21 576 Interestingly, this is reminiscent of the synchrony put forward by Attina *et al.* (2006) between  
22  
23 577 the hand target and the achievement of the consonantal closure in cued speech. This manually-  
24  
25 578 augmented speech communication system involves hand positions and finger configurations  
26  
27 579 added to the speech flow at a syllabic rhythm providing complementary information for deaf  
28  
29 580 people to almost perfectly understand oral communication. In their study of cued speech  
30  
31 581 production by French speakers, Attina and colleagues found a clear synchrony between the  
32  
33 582 apices of hand movements achieving their adequate position for a given syllable, and the  
34  
35 583 consonantal closure times in the corresponding speech.  
36

37 584

38 585 In summary, “the jaw meets the hand” in the sense that the jaw adapts the onset of its  
39  
40 586 movement in order for the pointing plateau to occur within the jaw cycle for the stressed  
41  
42  
43 587 syllable. This adaptation depends on the specific pointing and speech requirements. It occurs  
44  
45 588 within a temporal window providing two possible meeting points for the pointing gesture apex:  
46  
47 589 one corresponding to the apex of the first jaw opening gesture (when the first syllable is  
48  
49 590 stressed) and another corresponding to the onset of the second jaw opening gesture (when the  
50  
51 591 second syllable is stressed). The adaptation is achieved through the adequate delay of the onset  
52  
53 592 of the jaw movement. If the target to point at is far, the pointing stroke is longer by about 16 ms  
54  
55  
56  
57  
58  
59  
60

593 and the onset of the jaw movement is delayed by about 10 ms.

594

### 595 The hand meets the jaw

596 There is also a clear hand/finger adaptation to the movement of the jaw. It appears that the  
597 hand waits for the apex of the jaw opening stroke corresponding to the stressed vowel to  
598 occur before initiating its return stroke. This is illustrated by the observed significant effect of  
599 stress position on the pointing plateau offset. To our knowledge, this is the first experimental  
600 evidence of such adaptation of the pointing gesture to speech. This could be the way the hand  
601 “corrects” for the lack of jaw-finger synchrony at the apex time that is by providing another  
602 meeting point thanks to the expansion of the plateau duration. This ensures that the jaw  
603 opening gesture apex corresponding to the stressed syllable does occur within the pointing  
604 plateau in all cases.

605

### 606 Two coupled dynamic systems with synchronization sites

607 Altogether there appears to be a bidirectional link between the hand/finger and the jaw. This  
608 could be implemented by a series of local adaptations from one to the other, driven by  
609 specific temporal commands. The fact that there are simultaneous specific requirements for  
610 both systems rather suggests a kind of coupling between two dynamic systems, with the stress  
611 command specifying the adequate sites of synchronization. Actually, in the debate about  
612 modularity vs. interactivity of the motor control of speech and hand-pointing systems, it has  
613 been assumed that when the speech and the hand systems move at the same time, the brain  
614 would coordinate them as a single structure. In this framework, Kelso *et al.* (1981) showed  
615 that rhythmic tasks involving speech and hand movements displayed preferential phasing  
616 relationships between the articulators and the hand, as two coupled oscillator systems. Hence,  
617 according to these authors, speech-hand coordination should be described inside a coupled-

1  
2  
3  
4 618 oscillators modeling framework, such as the “Haken-Kelso-Bunz” (HKB) model introduced  
5  
6 619 by Haken *et al.* (1985) for bimanual coordination. Other contributions to the dynamic system  
7  
8 620 approach of speech-hand coordination showed that the link between the two systems is not  
9  
10 621 absolute (Smith *et al.*, 1986) and also depends on high-level cognitive factors (see Treffner  
11  
12 622 and Peter, 2002, for an improvement of the HKB model using intentional and attentional  
13  
14 623 factors). However, all these contributions support the view that the two systems could be  
15  
16 624 linked through their relative oscillatory frequencies. In addition, developmental studies  
17  
18 625 suggested a ratio of 2 to 1 between the preferential oscillatory frequencies of the jaw and the  
19  
20 626 hand performing pointing gestures (Ducey-Kaufmann *et al.*, in press, Ducey-Kaufmann,  
21  
22 627 2007). This implies that at most two syllables could be achieved within one pointing gesture  
23  
24 628 without influencing the pointing gesture. In a preliminary study (Rochet-Capellan *et al.*,  
25  
26 629 2007b), we showed that for a task consisting of pointing and naming a target with a 1-, 2-, 3-  
27  
28 630 vs. 4-CV syllable word twice in rapid succession (e.g. /pa/+gesture and /pa/+gesture again),  
29  
30 631 the delay between the apices of the two pointing gestures is the same for 1- and 2-CV words  
31  
32 632 while it significantly increases from 2- to 3-CV words. This preliminary result provides some  
33  
34 633 kind of confirmation of the 2 to 1 preferential ratio.

35  
36 634  
37  
38 635 Finally, the systematic delay of the speech gesture relative to the pointing gesture, together  
39  
40 636 with the large adaptation of speech to the hand gesture, suggest that the speech response could  
41  
42 637 be anchored in the pointing gesture through the online monitoring of the hand movement.  
43  
44 638 This could explain why Levelt *et al.* (1985) observed an effect of pointing perturbation on  
45  
46 639 speech only if the perturbation occurred at the beginning of the hand gesture. Thus, the speech  
47  
48 640 response would be carried by the pointing gesture, in agreement with the proposal that the  
49  
50 641 manual activity drives the coordination with the oral system (Iverson and Thelen, 1999). This  
51  
52 642 hypothesis would need further investigation, especially about the question of the information  
53  
54  
55  
56  
57  
58  
59  
60

643 used by the speech system to initiate its response.

644

## 645 **Conclusion**

646 Taken together, the results agree with the idea that speech focus is anchored in the pointing  
647 gesture for deictic expressions. This anchoring seems to be supported by a synchronization of  
648 the *speech-frame* (the jaw cycle) with the *sign-frame* (the pointing gesture) as suggested in  
649 the “Vocalize to Localize” framework for the origins of language. This results from two  
650 independent levels of speech adaptation. First, the effects of stress position and consonant  
651 order show an “internal” adaptation of the jaw opening gestures to the phonetic goal. In  
652 addition, the effect of spatial target position suggests an “external” adaptation for the  
653 synchronization with the pointing gesture within the adequate deictic site.

654

655 Further investigation will be needed to better understand the coordination of the jaw and hand  
656 in deictic tasks. But the interesting point about this study is that there *is* a coordination,  
657 implemented in a more or less complex way in the speaker’s brain, and that this coordination  
658 can be put forward and studied by analyzing online behavior in various linguistic tasks  
659 involving pointing using the voice and the hand.

660

## 661 **Acknowledgement**

662 This work is part of the “Patipapa” project funded by the French Ministry of Research (Action  
663 Concertée Incitative “Systèmes Complexes en Sciences Humaines et Sociales”). Rafael  
664 Laboissière works as a researcher at CNRS. We also wish to thank Christian Abry and  
665 Coriandre Vilain for insightful discussions, Marion Dohen for proofreading this manuscript,  
666 and the two anonymous reviewers for their helpful comments.

667 **References**

- 668 Abry, C., Vilain, A., and Schwartz, J. (2004). Introduction : Vocalize to Localize? A call for  
669 better crosstalk between auditory and visual communication systems researchers : From  
670 meerkats to humans. *Interaction Studies*, 5(3): 313-325.
- 671 Attina, V., Cathiard, M.-A., and Beautemps, D. (2006). Temporal measures of hand and  
672 speech coordination during French Cued Speech production, In: S. Gibet, N. Courty, & J.-  
673 F. Kamp (Eds.): *GW 2005, Lecture Notes in Artificial Intelligence*, vol. 3881, (pp. 13-24).  
674 Springer-Verlag.
- 675 Butterworth, G. (2003). Pointing is the royal road to language for babies. In: S. Kita (Ed.),  
676 *Pointing: Where Language, Culture, and Cognition meet* (pp. 9–33). Lawrence Erlbaum  
677 Associates.
- 678 Castiello, U., Paulignan, Y., and Jeannerod, M. (1991). Temporal dissociation of motor  
679 responses and subjective awareness. A study in normal subjects. *Brain*, 114 (Pt 6)(NIL):  
680 2639–55.
- 681 Diessel, H. (1999). Demonstratives: Form, Function, and Grammaticalization. *Typological*  
682 *Studies in Language*, 42, John Benjamins.
- 683 Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cognitive*  
684 *Linguistics* 17: 463-489.
- 685 Ducey-Kaufmann, V. (2007). *Le cadre de la parole et le cadre du signe : un rendez-vous*  
686 *développementale (The speech frame and the sign frame: a developmental rendez-vous)*,  
687 PhD thesis, Sciences du langage, Université Stendhal, Institut National Polytechnique,  
688 Grenoble.
- 689 Ducey-Kaufmann, V., Abry, C., and Vilain, C. (In press). When the speech frame meets the  
690 sign frame in a developmental framework. In *Emergence of Language Abilities :*  
691 *Ontogeny and phylogeny*, Lyon, France. ELA.
- 692 Feyereisen, P. (1997). The competition between gesture and speech production in dual-task  
693 paradigms. *Journal of Memory and Language*, 36(1): 13–33.

- 694 Gentilucci, M. (2003). Grasp observation influences speech production. *European Journal*  
695 *of Neuroscience*, 17: 179-184.
- 696 Gentilucci, M., Benuzzi, F., Gangitano, M., Grimaldi, S. (2001). Grasp with hand and  
697 mouth: a kinematic study on healthy subjects. *Journal of Neurophysiology*. 86: 1685-  
698 1699.
- 699 Gentilucci, M., Santunione, P., Roy, A.C., Stefanini, S. (2004). Execution and observation of  
700 bringing a fruit to the mouth affect syllable pronunciation. *European Journal of*  
701 *Neuroscience*, 19: 190-202.
- 702 Goldin-Meadow, S. and Butcher, C. (2003). Pointing toward two-word speech in young  
703 children. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 85-  
704 107). Lawrence Erlbaum Associates.
- 705 Green, J., Moore, C., and Reilly, K. (2002). The sequential development of jaw and lip  
706 control for speech. *Journal of Speech, Language, and Hearing Research*, 45: 66–79.
- 707 Haken, H., Kelso, J. A. S., and Bunz, H. (1985). A theoretical model of phase transitions in  
708 human hand movements. *Biological Cybernetic*, 51(5): 347–356
- 709 Haviland, J. B. (2000). Pointing, gesture spaces, and mental maps. In D. McNeill (Ed.),  
710 *Language and gesture* (pp. 13–46).
- 711 Holender, D. (1980). Interference between a vocal and a manual response to the same  
712 stimulus. In G. Stelmach and J. Requin (Eds.), *Tutorials in Motor Behavior* (pp. 421–432).  
713 North Holland Publishing Company.
- 714 Iverson, J. and Goldin-Meadow, S. (1998). Why people gesture as they speak. *Nature*,  
715 396:228.
- 716 Iverson, J. and Thelen, E. (1999). Hand, mouth, and brain : The dynamic emergence of speech  
717 and gesture. *Journal of Consciousness Studies*, 6: 19–40.
- 718 Kelso, J. A. D., Tuller, B. and Harris, K. S. (1981). A “Dynamic Pattern” perspective on the  
719 control and coordination of movement. In P.F. MacNeilage (Ed.), *The production of*  
720 *speech* (pp. 137-173). New York: Springer-Verlag.

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
- 721 Levelt, W. J. M., Richardson, G. and Heij, W. L. (1985). Pointing and voicing in deictic  
722 expressions. *Journal of Memory and Language*, 24: 133–164.
- 723 Løevenbruck, H., Baciú, M., Segebarth, C. and Abry, C. (2005). The left inferior frontal gyrus  
724 under focus: an fMRI study of the production of deixis via syntactic extraction and  
725 prosodic focus. *Journal of Neurolinguistics*, 18: 237–258.
- 726 MacNeilage, P. and Davis, B. (2000). On the origins of internal structure of word forms.  
727 *Science*, 288: 527–531.
- 728 McNeill, D. (1981). Action, thought and language. *Cognition*, (10) :201–208.
- 729 McNeill, D. (1992). *Hand and Mind*. Chicago: University of Chicago Press.
- 730 McNeill, D. (2000). Introduction. In D. McNeill (Ed.), *Language and gesture* (pp. 1–10).
- 731 Munhall, K. and Jones, J. (1998). Articulatory evidence for syllabic structure. *Behavioral and*  
732 *Brain Science*, 21: 524–525.
- 733 Petitto, L., Holowka, S., Sergio, L. and Ostry, D. (2001). Language rhythms in baby hand  
734 movement. *Nature*, 413(6851): 35–36.
- 735 Pizzuto, E., Capobianco, M. and Devescovi, A. (2005). Gestural-vocal deixis and  
736 representational skills in early language development. *Interaction Studies. Social*  
737 *Behaviour and Communication in Biological and Artificial Systems*, 6(2): 223–252.
- 738 Rochet-Capellan, A. and Schwartz J-L. (2007a). An articulatory basis for the Labial-to-  
739 Coronal effect: /pata/ seems a more stable articulatory pattern than /tapa/. *Journal of*  
740 *Acoustic Society of America*, 121(6): 3740-3754.
- 741 Rochet-Capellan, A., Schwartz J-L., Laboissière, R. and Galván, A. (2007b). Two CV  
742 syllables for one pointing gesture as an optimal ratio for jaw-arm coordination in a deictic  
743 task: A preliminary study. In *proceedings of EuroCogSci07, the European cognitive*  
744 *Science conference*, pp. 608-613, Delphi, Greece.
- 745 Stetson, R. H. (1951). *Motor Phonetics* (North-Holland, Amsterdam), 2nd ed.
- 746 Smith, A., McFarland, D. H. and Weber, C. M. (1986). Interactions between speech and  
747 finger movements: an exploration on the dynamic pattern perspective. *Journal of Speech*

1  
2  
3 748 *and Hearing Research*, 29(4): 471–80.

4  
5  
6 749 Tomasello, M., Carpenter, M. and Liszkowski U. (2007). A new look at infant pointing. *Child*  
7  
8 750 *Development*, 78(3):705-722.

9  
10 751 Treffner, P. and Peter, M. (2002). Intentional and attentional dynamics of speech-hand  
11  
12 752 coordination. *Human Movement Science*, 21(5-6): 641–97.

13  
14  
15 753 Volterra, V., Caselli, M. C., Capirci, O., and Pizzuto, E. (2005). Gesture and the emergence  
16  
17 754 and development of language. In M. Tomasello and D. Slobin (Eds.), *Beyond nature-*  
18  
19 755 *nurture- Essays in honor of Elizabeth Bates* (pp. 3–40). Mahwah, N. J.: Lawrence Erlbaum  
20  
21 756 Associates.

22  
23 757 Wilkins, D. (2003). Why pointing with the index finger is not a universal (in sociocultural and  
24  
25 758 semiotic terms). In S. Kita (Ed.), *Pointing: Where language culture and cognition meet*,  
26  
27 759 (pp. 171-215). Lawrence Erlbaum Associates.

28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
760

761 **Tables**762 Table 1:

763 Number of rejected trials over all 20 participants and all experimental conditions: mean  
764 number, standard error (Se), minimum and maximum number of utterance errors (incorrect  
765 pronunciation), initiation errors (response before the go-signal), correct trials for which the  
766 two finger IREDs were partially masked and for which the jaw IRED was partially masked,  
767 see text for detail.

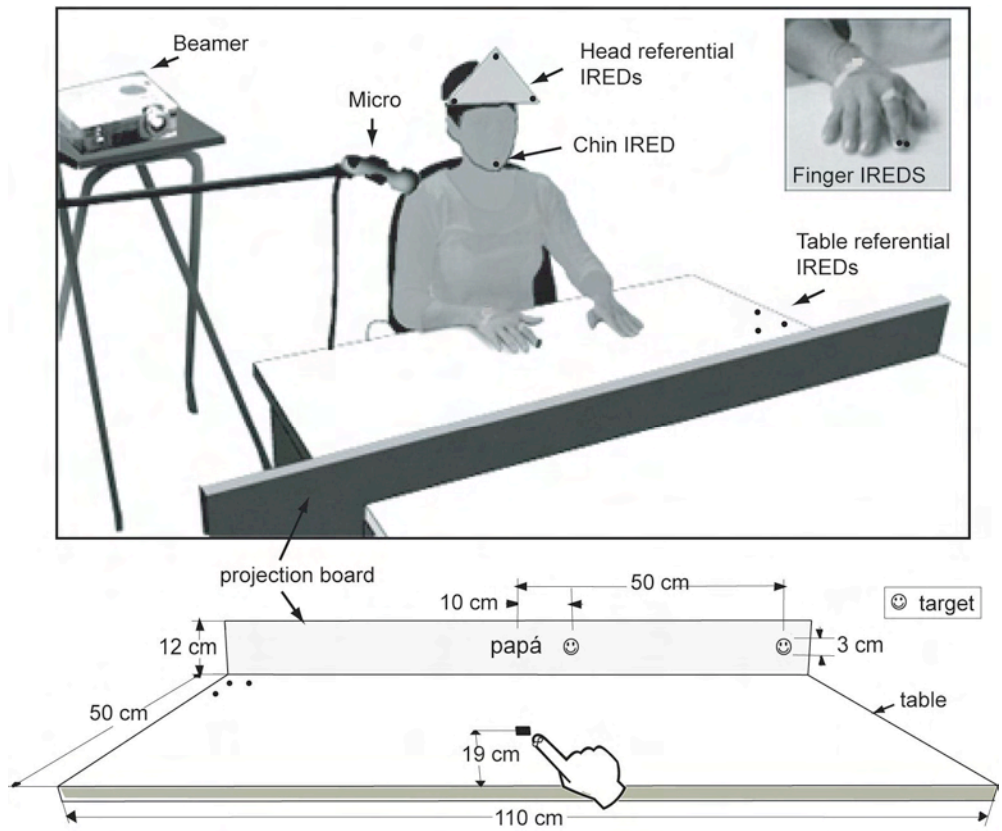
768

769

	<b>Mean</b>	<b>Se</b>	<b>Min</b>	<b>Max</b>
<b>Utterance error</b>	4.1	0.86	0	11
<b>Initiation error</b>	2.35	1.94	0	39
<b>Masked finger</b>	8.2	1.50	0	21
<b>Masked jaw</b>	0.75	0.60	0	12

770

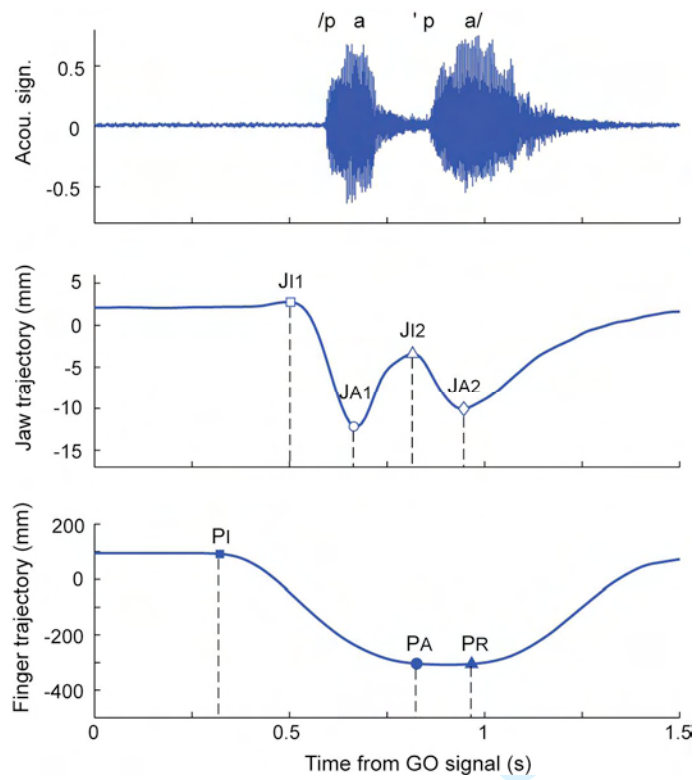
771 **Figures**



772

773 Figure 1

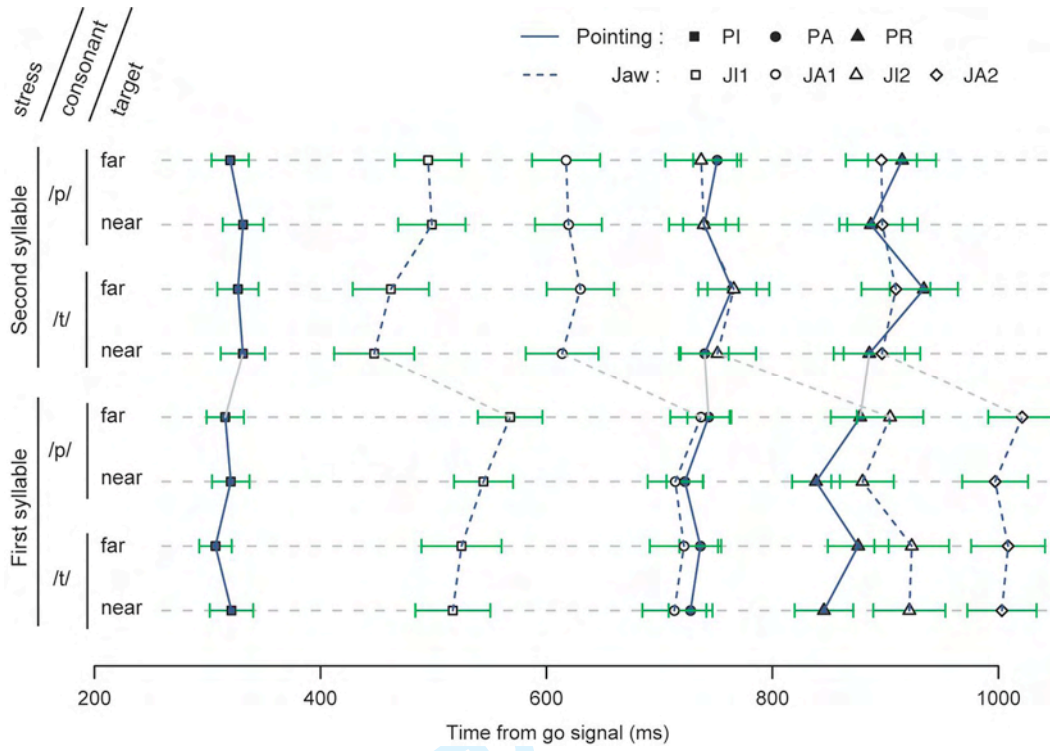
review



774

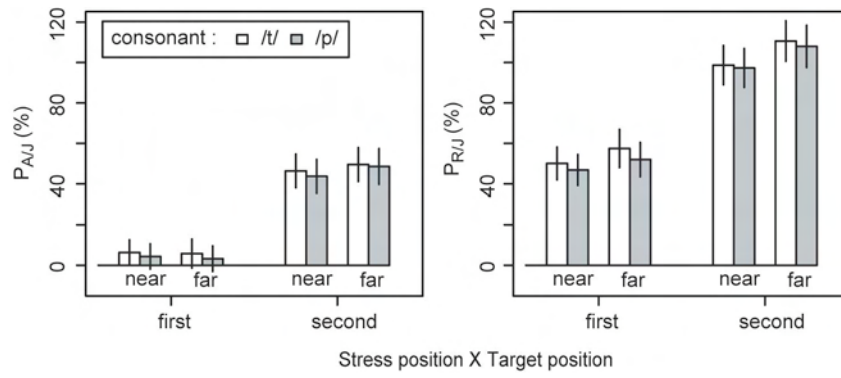
775 Figure 2

Peer Review



776

777 Figure 3



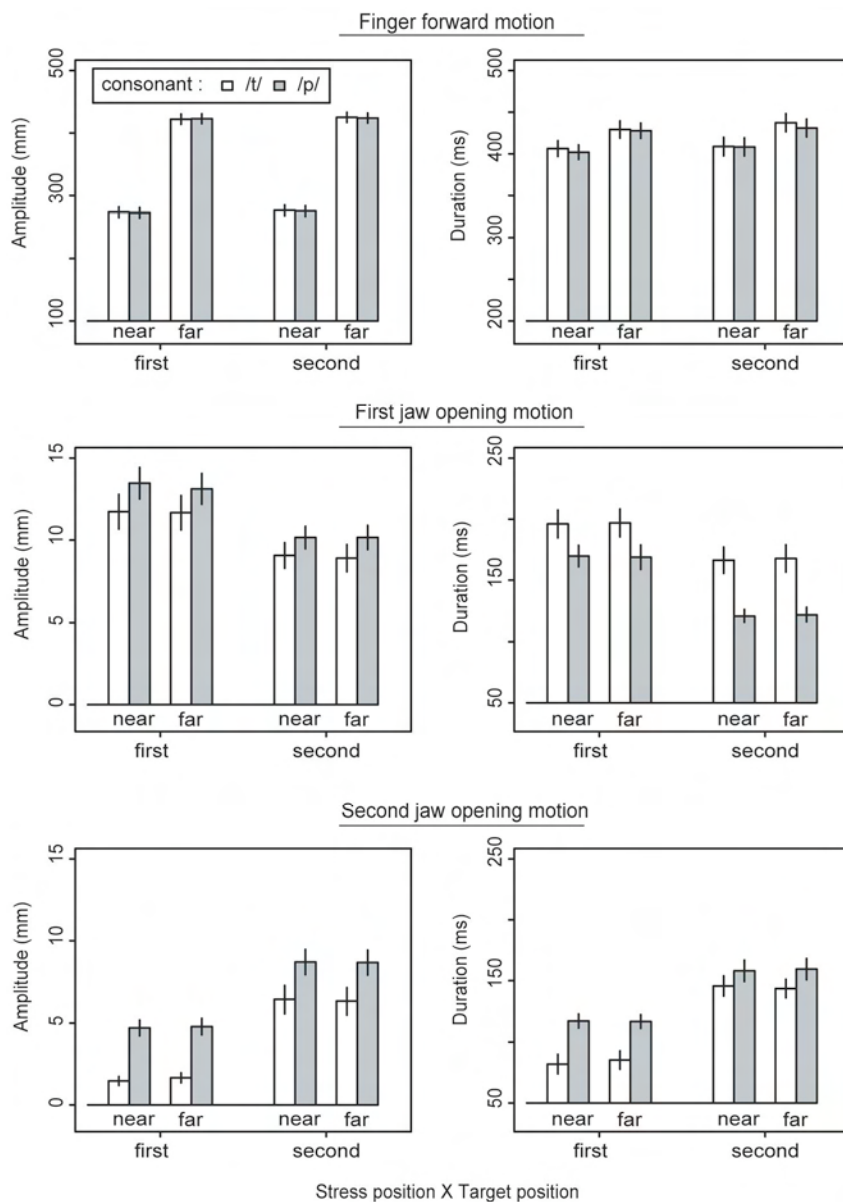
778

779

780 Figure 4

For Peer Review

781



782

783 Figure 5

## 784 **Figures captions**

### 785 Figure 1:

786 Experimental setup (top) and projection dimensions (bottom). The participants pointed at, and  
787 named a smiley sign projected on a screen in front of them. Finger and jaw movements were  
788 captured using Optotrak (in front of the participant) with two IREDs for the finger, one for the  
789 jaw (chin IRED), three for the head (head IRED referential), and three on the table (referential  
790 for all moving IREDs). The target name appeared in the midline of the visual field (e.g.  
791 /pa'pa/) and the pointing target on the right side, either at a near (10 cm) or a far (50 cm)  
792 position.

### 794 Figure 2:

795 Example of signals and labeling for a /pa'pa/ trial. Top panel: acoustic signal; middle panel:  
796 jaw trajectory; bottom panel: finger trajectory. Onsets and offsets of strokes are defined as the  
797 instant when velocity reaches 10% of the maximum value during the corresponding stroke.  
798  $J_{11}$ ,  $J_{12}$  and  $J_{A1}$ ,  $J_{A2}$  respectively correspond to the onset and apex (offset) times of the first and  
799 second jaw opening gestures.  $P_I$  and  $P_A$  correspond to the onset and apex (offset) times of the  
800 pointing forward stroke.  $P_R$  is the onset time of the finger return movement. Note that, in this  
801 example, even though stress position affects the second syllable, the amplitude of the jaw  
802 movement is smaller for the second syllable than for the first one. This corresponds to the  
803 general pattern observed in this experiment (see jaw opening movement panels in Figure 5).

### 805 Figure 3:

806 Means of the elapsed time from the go signal to the pointing and jaw events relative to stress  
807 position (first vs. second syllable), consonant (/t/ vs. /p/) and target position (near vs. far).  $P_I$ ,  
808  $P_A$  and  $P_R$  respectively correspond to the onset, the apex and the onset of the return of the

1  
2  
3 809 finger pointing movement.  $J_{I1}$ ,  $J_{A1}$  and  $J_{I2}$ ,  $J_{A2}$  respectively correspond to the onset and the  
4  
5 810 apex of the first and the second jaw opening strokes.  
6  
7

8 811

9  
10 812 Figure 4.

11  
12 813 Means and standard errors of the time of the pointing apex ( $P_{A/J}$ , left) and of the time of the  
13  
14 814 onset of the pointing return ( $P_{R/J}$ , right) relative to the apices of the two jaw opening gestures,  
15  
16 815 depending on stress position (first vs. second syllable), target position (near vs. far) and  
17  
18 816 consonant (/t/ vs. /p/).  
19

20 817

21  
22 818 Figure 5.

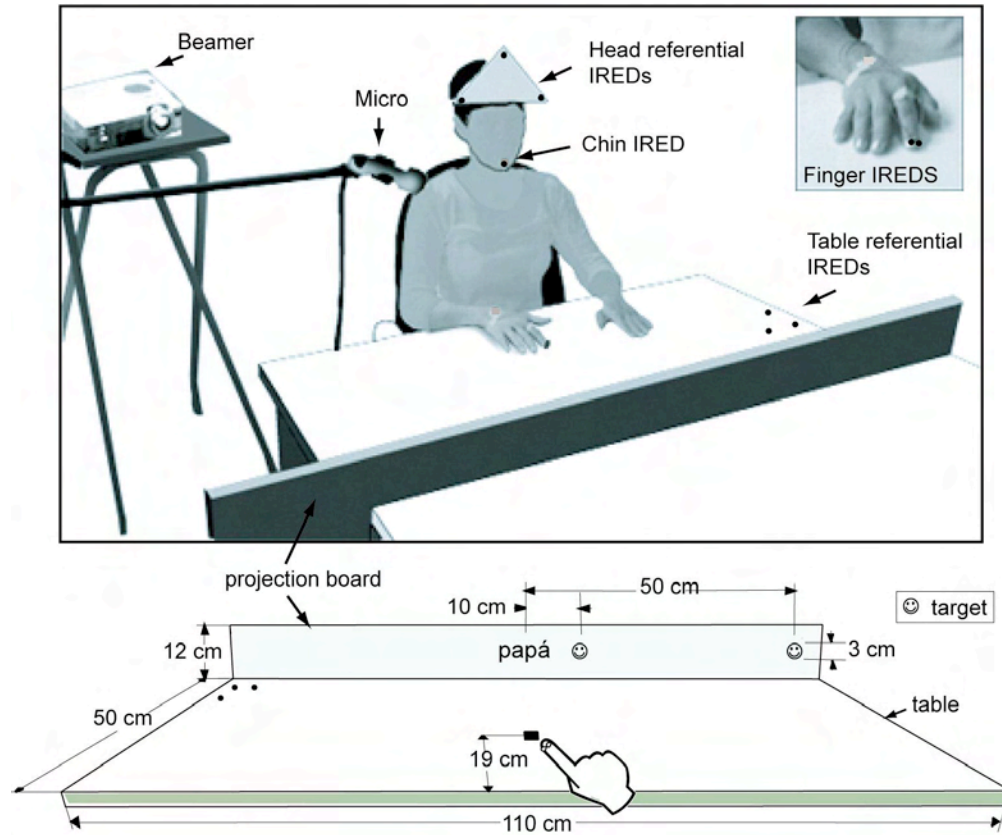
23  
24 819 Means and standard errors of the amplitude and the duration of the finger forward stroke (first  
25  
26 820 row) and of the two jaw opening strokes (second and third rows) depending on stress position  
27  
28 821 (first vs. second syllable), target position (near vs. far) and consonant (/t/ vs. /p/).  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Table 1:

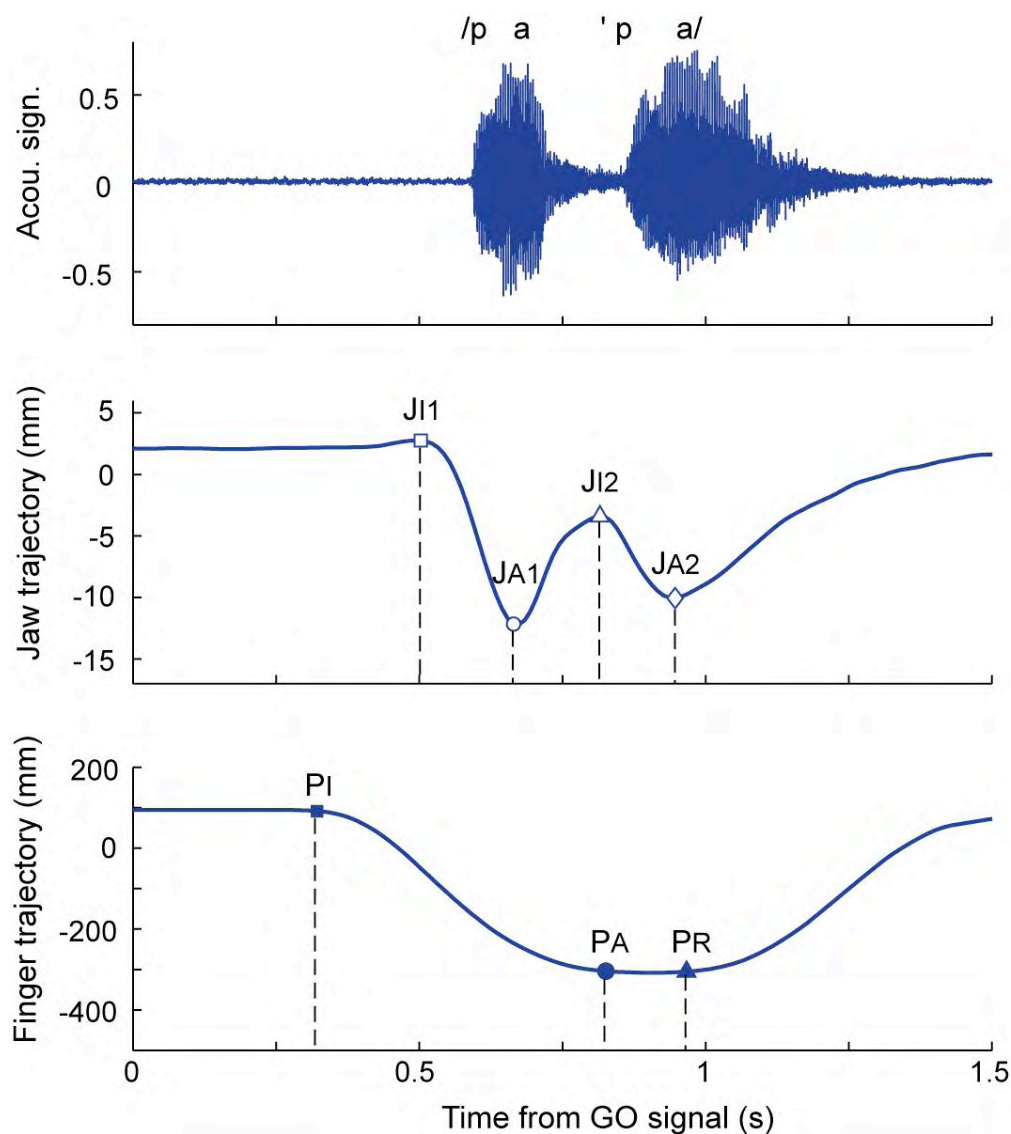
Number of rejected trials. For the 20 participants and all experimental conditions: mean number, standard error (Se), minimum and maximum number of: utterance errors (incorrect pronunciation); initiation errors (response before go-signal); correct trials with the two finger IREDs partially masked and with the jaw IRED partially masked, see text for detail.

	Mean	Se	Min	Max
<b>Utterance error</b>	4.1	0.86	0	11
<b>Initiation error</b>	2.35	1.94	0	39
<b>Masked finger</b>	8.2	1.50	0	21
<b>Masked jaw</b>	0.75	0.60	0	12



**Figure 1: Experimental setup (top) and projection dimensions (bottom). The participants pointed at, and named a smiley sign projected on a screen in front of them. Finger and jaw movements were captured using Optotrak (in front of the participant) with two IREDs for the finger, one for the jaw (chin IRED), three for the head (head IRED referential), and three on the table (referential for all moving IREDs). The target name appeared in the midline of the visual field (e.g. /pa'pa/) and the pointing target on the right side, either at a near (10 cm) or a far (50 cm) position.**

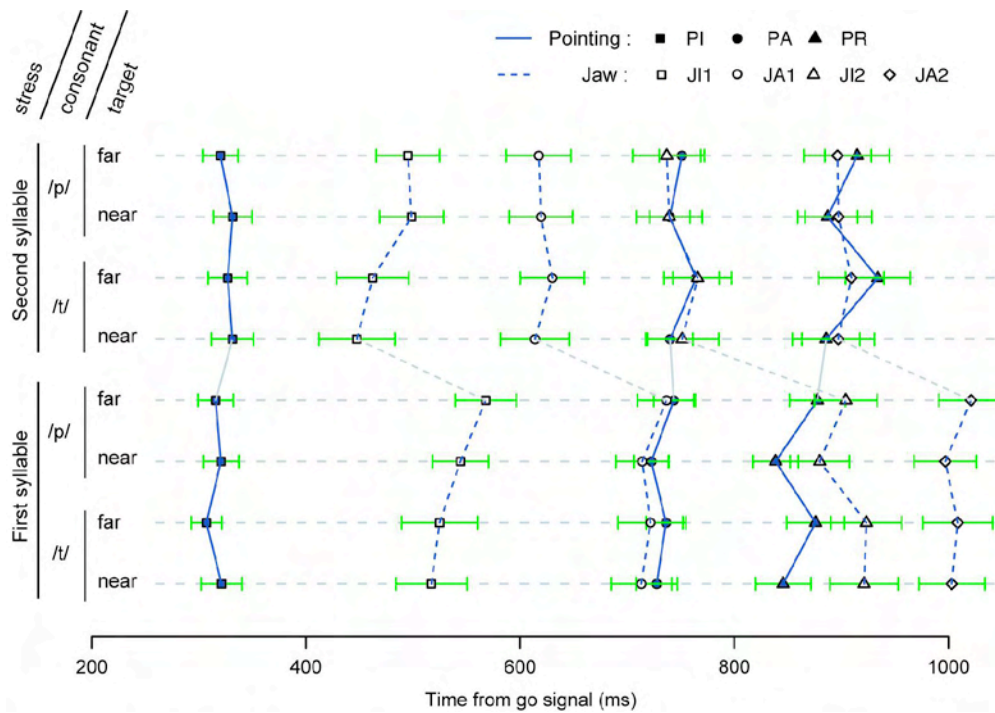
130x111mm (300 x 300 DPI)



**Figure 2: Example of signals and labeling for a /pa'pa/ trial. Top panel: acoustic signal; middle panel: jaw trajectory; bottom panel: finger trajectory. Onsets and offsets of strokes are defined as the instant when velocity reaches 10% of the maximum value during the corresponding stroke. JI1, JI2 and JA1, JA2 respectively correspond to the onset and apex (offset) times of the first and second jaw opening gestures. PI and PA correspond to the onset and apex (offset) times of the pointing forward stroke. PR is the onset time of the finger return movement. Note that, in this example, even though stress position affects the second syllable, the amplitude of the jaw movement is smaller for the second syllable than for the first one. This corresponds to the general pattern observed in this experiment (see jaw opening movement panels in Figure 5).**

90x101mm (300 x 300 DPI)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

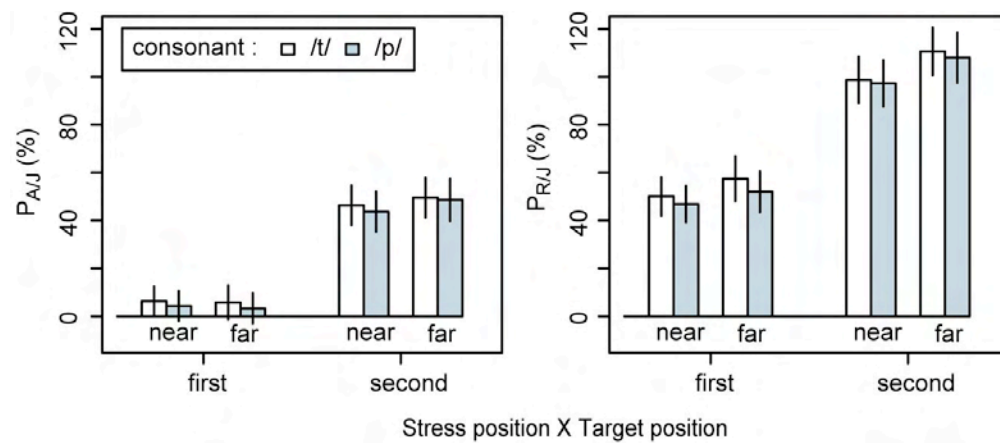


**Figure 3: Means of the elapsed time from the go signal to the pointing and jaw events relative to stress position (first vs. second syllable), consonant (/t/ vs. /p/) and target position (near vs. far). PI, PA and PR respectively correspond to the onset, the apex and the onset of the return of the finger pointing movement. JI1, JA1 and JI2, JA2 respectively correspond to the onset and the apex of the first and the second jaw opening strokes.**

139x97mm (300 x 300 DPI)

view

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

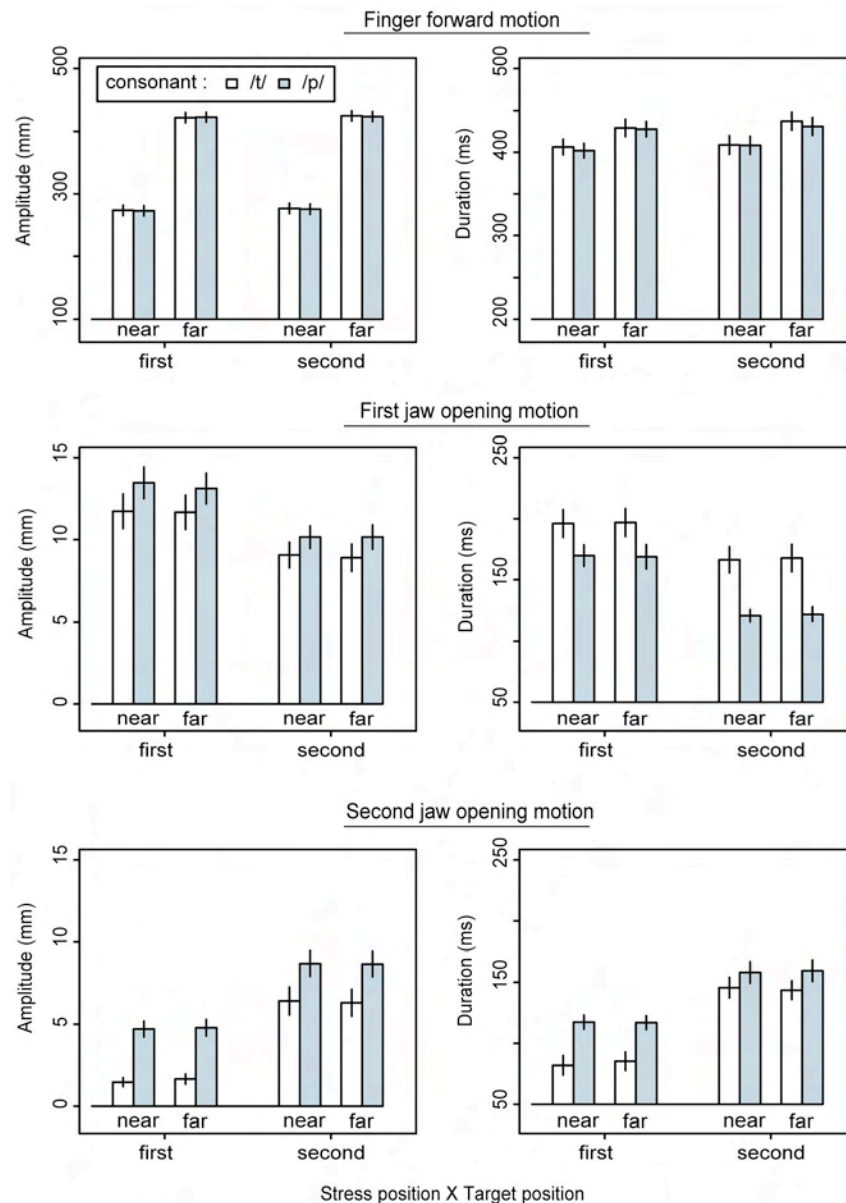


**Figure 4. Means and standard errors of the time of the pointing apex (PA/J, left) and of the time of the onset of the pointing return (PR/J, right) relative to the apices of the two jaw opening gestures, depending on stress position (first vs. second syllable), target position (near vs. far) and consonant (/t/ vs. /p/).**

111x48mm (300 x 300 DPI)

Peer Review

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



**Figure 5. Means and standard errors of the amplitude and the duration of the finger forward stroke (first row) and of the two jaw opening strokes (second and third rows) depending on stress position (first vs. second syllable), target position (near vs. far) and consonant (/t/ vs. /p/).**

111x158mm (300 x 300 DPI)