Finger-jaw coordination during a deictic gesture with CVCV utterances: the effect of stress position

Amélie Rochet-Capellan, Jean-Luc Schwartz, Rafael Laboissière, Arthuro Galván

Finger-jaw coordination during a deictic gesture with CVCV utterances: the effect of stress position

Amélie Rochet-Capellan[1], Jean-Luc Schwartz[1], Rafael Laboissière[2], Arturo Galván[2]

[12]Institut de la Communication Parlée UMR CNRS 5009 - INP Grenoble
46, avenue Félix Viallet 38031 GRENOBLE CEDEX 1, France

[34] Max Planck Institute for Human Cognitive and Brain Sciences,
Amalienstrasse 33 80799 Munich, Germany

amelie.rochet-capellan@icp.inpg.fr, jean-Luc.schwartz@icp.inpg.fr,
laboissiere@cbs.mpg.de, galvan@cbs.mpg.de

Abstract. *The present study investigates the hypothesis of a synchronization between the extremum (apex) of the pointing gesture and the apex of the jaw opening gesture for the stress vowel in 'CVCV vs CV'CV utterances in a deictic task (consisting in showing a target while naming it). The results show that for 'CVCV sequences (e.g. /pápa/) the pointing apex tends to be phased with the jaw apex for the first vowel. For CV'CV (e.g. /papá/), the pointing apex happens more or less at the mid-point between the jaw apices for the first and the second vowels. However, the onset of the hand return movement is synchronized with the jaw apex for the second vowel. The absolute timing of the forward movement of the finger does not depend on the stress condition, indicating that the differences in the relative timing between the jaw and the hand across conditions is a consequence of the speech gesture being adapted to the hand gesture. These results are discussed in relation to the results of Levelt et al. (1985) on speech-hand coordination in deictic expressions and regard to Abry et al.'s (2004) "Vocalize to Localize" framework.*

1. Introduction

1.1. General framework and objectives

The deictic gesture has the particularity to require the synchronization between speech and hand. Moreover, for many theorists, it constitutes the primary indexical sign in both phylogeny and ontogeny (Haviland, 2000). According to the "Vocalize to Localize" framework (Abry et al., 2004), the association of voice and hand in deictic tasks is a key step of both phylogenetic and ontogenetic developments of oral communication. Their hypothesis is that the deictic function is first associated with hand movements and is later connected to the speech. This leads to the close association between the pointing gesture and the "foot", the coordinative unit that gives the basis of the speech rhythm. Two consequences could be derived from this hypothesis. First, hand-voice coordination might occur around the foot stress unit. Second, this coordination might be constrained by the physical and motor properties of the speech and the arm-hand-finger systems. In addition, motor control studies (Nelson, 1984) and speech production theories (MacNeilage, 1998) suggest that the jaw plays a basic role in the control of the speech rhythm. Therefore, there should be a close coordination between jaw and hand in deictic tasks.

1.2. Investigation of synchronization processes in deictic expressions

Levelt et al. (1985) investigated the processes underlying the synchronization between voicing and pointing in deictic expressions. The focus was on Gesture plus Speech (GS) task, associating a hand pointing gesture with a verbal response ("dit lampje" vs. "dat lampje", i.e. "this lamp" vs. "that lamp"). Hand and speech performances in the GS task were respectively compared with hand performances in a gesture only (G) task and speech performances in a speech only (S) task. Their main results are: (1) The presence of speech does not affect the time of the pointing apex and has a small delay effect on the initiation time of pointing. (2) On the contrary, the presence of the hand gesture delays the voice onset in a way that tends to synchronize the voice onset with the pointing apex. (3) A mechanical perturbation applied to the hand gesture affects the voice onset only if it is applied during the first stage of motion. According to the authors, these results show that the interaction between the two systems is limited to the planning stage as well as to the initial portion of the execution stage. Moreover, as Holender (1980) and Feyreisen (1997), they consider this interaction as a competition for resources access.

1.3. Investigation of synchronization sites

The precedent studies did not explore the possible *sites* of synchronization between the orofacial and the hand gestures. The Levelt et al.'s study provides a first answer to this question: voice onset tends to be synchronized with hand pointing apex. However, Levelt et al. did not vary the position of the speech deictic site, always at the beginning of the utterance ("dat" vs. "dit"). The explanation of their results in the "Vocalize to Localize" framework would be that the speech indexical site tends to be synchronized with the hand pointing apex. Consequently, the speech-hand relationship would go beyond the competition for the common resources during the planning phase. Moreover, in Levelt et al. (1985) speech timing is measured only by the voice onset. One could argue that mandible motion measurements would be more appropriate for studying speech-hand coordination. In this paper, we propose to study how the variation of the stress position in CVCV words affects the coordination between the jaw and the hand in a GS pointing task. Our main hypothesis is that the hand pointing apex would be synchronized with the apex of the jaw opening gesture for the stressed vowel, that is the first vowel in the case of 'CVCV utterances and the second vowel in the case of CV'CV utterances. This alignment could be reached either only by jaw adaptation, the hand motion remaining the same for both 'CVCV and CV'CV sequences, or by a mutual adaptation, involving a modification of both jaw and hand motions from 'CVCV to CV'CV.

2. Method

Speakers. Twenty native Brazilian Portuguese speakers (4 men and 16 women) participated in the experiment. They were all right-handed, with no speech or hearing pathology record and unaware of the purpose of the experiment. The Brazilian

Portuguese language was chosen because it is possible to find pairs of words in this language that differ only by the position of the stress.

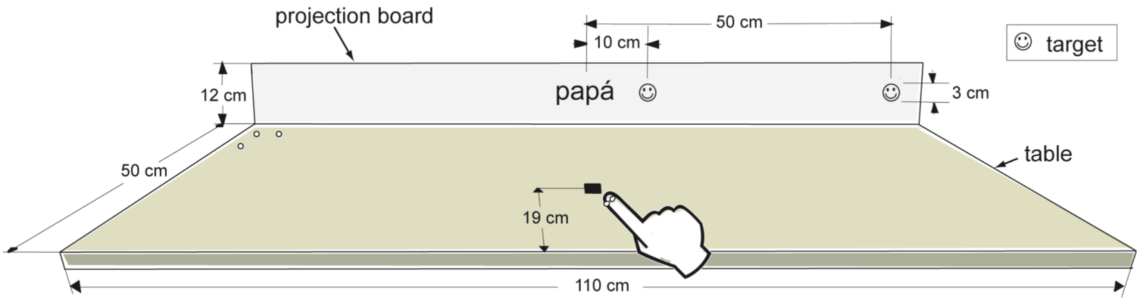

Figure 1. Experimental setup

Procedure. As in the princeps study by Levelt et al. (1985), our experiment involved a hand-pointing task associated with a verbal response. The main factor was the stress position in the CVCV sequence (stress on the first vs. the second vowel), which were either /papa/ or /tata/. The vowel /a/ was selected since its realization requires a large jaw opening gesture. Two spatial targets were used for the pointing gesture (near vs. far). Speakers were seated at a table. The targets to point and the item to pronounce were projected on a white board in front of the speaker using a beamer. A black square pasted on the midline of the table indicated the rest position of the finger. Speakers were informed that a word and a red smiley ☺ sign (the target) would appear on the board. The target appeared in the right visual field, either at the near or at the far position (see Fig. 1). In order to make the joint gesture/pronunciation task natural, the subjects were instructed that the displayed word was the name of the person represented by the smiley-target. As soon as the sign color changed from red to green, the subject should point to the smiley and name it at the same time. To become used with the task, subjects were briefly trained prior to the experiment start to point to objects in the room while naming them. Also, the subjects trained to read the CVCV sequences aloud in order to ensure that they understood the stress instruction. Finger and jaw movement were recorder using an Optotrak. Two Optotrak markers were pasted at the tip of index finger, such that at least one of them was always visible by the cameras during the course of the pointing movement. The jaw position was tracked by a third marker attached to the chin. Three other fixed sensors were pasted on the table to provide a referential for the moving sensors. Head motion was measured by three markers attached to a plastic triangle, which was fixed by a strap around the subject's head. Jaw positions were computed in relation to the head moving reference frame. The experiment was divided into four blocks. A block contained 4 practice trials and 40 experimental trials, 10 for each [stress position]*[target position] experimental condition with equal number of /papa/ and /tata/. The order of the trials was randomized for each block and each speaker. The blocks were separated by a 30 seconds pause. The presentation of the red target lasted for 3.5 s plus a Gaussian variation with zero mean and 0.15 s standard deviation. Then, the target became green (the go signal) and lasted 1 s on the board.

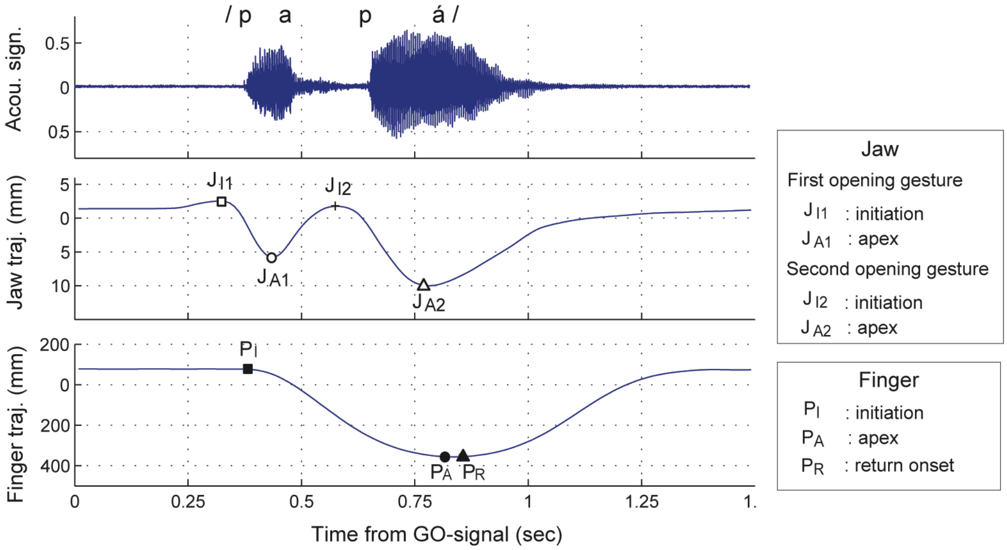

Figure 2. Example of record and labeling for a /papa/ trial.

Data recording and processing. The Optotrak markers positions were sampled against time at 99.474 Hz and the sound was recorded at 16 kHz with a microphone. In order to reduce the amount of data to process, we conducted a Principal Component Analysis (PCA) on the Optotrack data. The first PC explained most of the variance for all subjects: 98.8% (std = 0.7%) and 98.3% (std = 1.2%) for the two finger markers and 95.6% (std = 2.2%) for the jaw marker. Thus, the first PC appears as a good representation of finger and jaw movements. These signals were lowpass filtered at 15 Hz with a Butterworth filter. The temporal measurements are indicated in Fig. 2. Motion onset (initiation) and apex times correspond to 10% of the peak velocity at the beginning and the end of the movement, respectively. P_I and P_A are the initiation and apex times for the forward movement of the finger, while P_R is the onset of the return movement. For the jaw, J_{I1}, J_{A1} and J_{I2}, J_{A2} are the initiation and the apex times of the first and second opening gestures, respectively. Labeling errors were manually corrected.

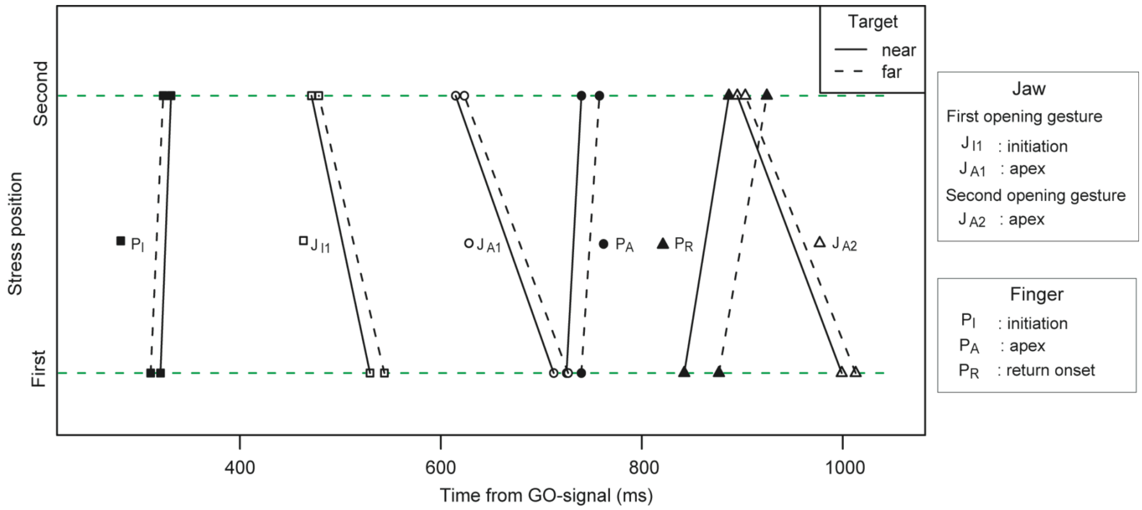

Figure 3. Mean durations from the GO-signal to various jaw and finger temporal events, as a function of stress and target position.

3. Results

3.1. Jaw-finger synchronization

The main results are shown in Fig. 3. P_A is close to J_{A1} in the first-stress condition whereas it is more or less at the mid-point between J_{A1} and J_{A2} in the second-stress condition. On the contrary, P_R is close to J_{A2} in the second-stress condition and about at the mid-point between J_{A1} and J_{A2} in the first-stress condition. We computed the relative finger/jaw synchronization indices for the apex and return instants of the finger as, respectively:

$$P_{A/J} = (P_A - J_{A1})/(J_{A2} - J_{A1}) \text{ and } P_{R/J} = (P_R - J_{A1})/(J_{A2} - J_{A1})$$

Values of 0 and 1 for these indices indicate synchronization with the first and the second jaw apex, respectively. Within-subject ANOVA with stress and target position as factors on $P_{A/J}$ and $P_{R/J}$ confirms the previous pattern (see Fig. 4). Indeed, the $P_{A/J}$ mean is close to 0 (0.05) in the first-stress condition and to 0.5 (0.47) in the second-stress one ($F(1,19) = 65.2$, $p < .0001$), while the $P_{R/J}$ mean is close to 0.5 (0.52) in the first-stress condition and to 1 (1.04) in the second-stress one ($F(1,19) = 100.9$, $p < .0001$). The target position has no effect on $P_{A/J}$ (near: 0.25, far: 0.27, $F(1,19) = 1.3$, $p = 0.3$) while it affects $P_{R/J}$ (near: 0.73, far: 0.82, $F(1,19) = 11.9$, $p < .003$). The stress*target interaction is not significant for both $P_{A/J}$ and $P_{R/J}$ ($P_{A/J}$: $F(1,19) = 4.1$, $p = .058$; $P_{R/J}$: $F(1,19) = 4.1$, $p = .055$).

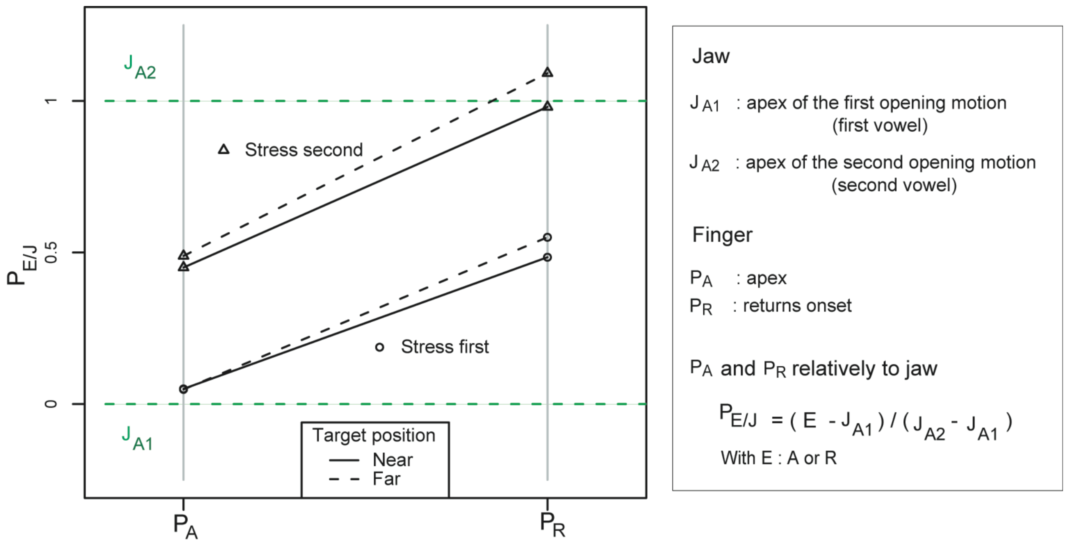

Figure 4. Means of P_A and P_R values relatively to ($J_{A2} - J_{A1}$) according to stress and target position.

3.2. Unidirectional or bi-directional adaptation?

Initiation times. From Fig. 3, we see that J_{I1} occurs after P_I. The mean ($J_{I1} - P_I$) value is greater in the first-stress condition (223 ms) than in the second-stress one (149 ms, $F(1,19) = 13.7$, $p < .002$). Yet, the difference between P_I values across stress conditions is not significant ($F(1,19) = 2.2$, $p = .08$). On the contrary, J_{I1} occurs 62 ms earlier in the second-stress (475 ms) condition than in the first-stress one (537 ms, $F(1,19) = 12.1$, $p < .0008$). In addition, the ($J_{I1} - P_I$) mean is higher in the far-target condition (195 ms) than in the near-target one (176 ms, $F(1,19) = 10.1$, $p < .005$).

Arrival times. The increase of target distance induces a significant16 ms delay for P_A (near: 733 ms, far: 749 ms, $F(1,19) = 9.2$, $p < .007$) and a significant 12 ms delay for both J_{A1} (near: 664 ms, far: 676 ms, $F(1,19) = 5.5$, $p < .03$) and J_{A2} (near: 946 ms, far: 958 ms, $F(1,19) = 6.6$, $p < .02$). The difference between P_A values in the first- (732 ms) and second-stress (749 ms) conditions is not significant ($F(1,19) = 3.0$, $p = .10$). On the contrary, J_{A1} and J_{A2} happen earlier in the second-stress condition than in the first-stress one (from 619 to 720 ms for J_{A1}, $F(1,19) = 58.4$, $p < .0001$ and from 899 to 1005 ms for J_{A2}, $F(1,19) = 59.7$, $p < .0001$).

Durations and amplitudes. Fig. 3 suggests that the elapsed time from P_I to P_A (D_P) increases with target distance (near: 406 ms, far: 431 ms, $F(1,19) = 26.3$, $p < .0001$). This comes together with an increase of the distance moved by the finger (A_P) (near: 275 mm, far: 423 mm, $F(1,19) = 2220.9$, $p < .0001$). On the contrary, stress position does not affect D_P ($F(1,19) = 1.4$, $p = .3$) while A_P tends to be larger in the second-stress (350 mm) condition as compared to the first-stress one (348 mm, $F = 8.3$, $p < .01$). The elapsed time from P_A to P_R (pointing plateau) is longer in the far-target condition (152 ms) than in the near one (132 ms, $F(1,19) = 13.8$, $p < .002$) and in the second-stress condition (157 ms) than in the first-stress one (127 ms, $F(1,19) = 14.0$, $p < .001$). Finally, duration (D_{J1}) and amplitude (A_{J1}) of the first jaw opening motion is higher in the first-stress condition than in the second one (table 1, +41 ms, $F(1,19) = 275$, $p < .0001$ and +2.9 mm, $F(1,19) = 20.5$, $p < .0002$). The reverse pattern characterizes the second gesture with greater D_{J1} and A_{J1} values in the second-stress condition than in the first one (+48 ms, $F(1,19) = 40.9$, $p < .0001$ and +3.8 mm, $F(1,19) = 40.1$, $p < .0001$). Target position affects neither A_{J1} and A_{J2} nor D_{J1} and D_{J2}.

Stress	**First**		**Second**	
Target	**Near**	**Far**	**Near**	**Far**
Pointing forward	273	422	276	424
First jaw opening motion	12.6	12.4	9.6	9.6
Second jaw opening motion	3.1	3.3	7.5	7.5

Table 1. Means of amplitude (mm) of finger forward and jaw opening motions according to stress and target position.

3. Discussion and conclusion

In agreement with our initial hypothesis, stress position affects jaw-finger coordination. In the first-stress condition, the finger apex P_A is synchronized with the first-syllable jaw apex J_{A1}. In the second-stress condition, the finger return instant P_R is synchronized with the second-syllable jaw apex J_{A2}. It seems then that the remarkable points of the finger trajectory tend to be anchored to the jaw apices of the stressed syllable. These results hold for both near and far target positions. The fact that the finger pointing movement is quite similar across the stress conditions indicates that the jaw movement adapts itself to the hand movement in order to achieve the synchronizations described above. This main result is in agreement with Levelt et al.'s (1985) results.

The position of the target, either near or far, affects both finger and jaw motions. For the finger, the duration of the movement is larger for the farther target position whereas the motion initiation time is unchanged. In addition, the duration of the pointing plateau is about 20 ms longer in the far-target condition in respect to in the near-target one. As regards the jaw movement, target position does not affect amplitude and duration of opening strokes but induces delays on the two apices values (increase of 12 ms for both J_{A1} and J_{A2}). Again, these results suggest that speech adapts to pointing gestures in deictic expressions.

As in Levelt et al.'s study, in which the stress is always placed in the first syllable of the utterance, we found that finger apex is synchronized with the jaw apex of the first syllable in the first-stress condition. However, the pattern of synchronization for the second-stress condition is puzzling. Indeed, when the second syllable is stressed, finger apex happens at about the mid-point between the two jaw apices (about 130 ms after J_{A1} and 150 ms before J_{A2}). The pointing apex is then phased neither with voice onset nor with the jaw apex for the stressed vowel. This pattern may result from two types of jaw adaptation: (1) adaptation to the phonological goal, with a reduction of the first opening motion when the first vowel is unstressed; (2) adaptation for speech-pointing coordination, with a lead of jaw initiation time in the second-stress condition as compared to the first-stress one. In addition, we observed that the finger waits for the second jaw apex before starting the return movement.

Results coming from speech development studies can shed some light on why the synchronization pattern is different across stress conditions. Observations of speech and gestures produced by infants (Ducey-Kaufmann et al., 2005) showed that the natural frequency of the jaw system is twice that of the hand system. This 2:1 ratio of natural frequencies may be rooted on biomechanical constraints of both the upper limb and the orofacial systems. In order to take this into account, the Central Nervous System would try to accommodate two opening/closing movements of the jaw into one forward/backward movement of the finger. For this aim, one possible strategy would be to always synchronize the finger apex with the jaw apex of the first syllable, regardless of the stress placement. This is what happens in the first-stress condition, but not in the second-stress one. A possible explanation would be that, on the top of this 2:1 timing ratio restriction, there is a communication constraint related to the vocal site of the deictic expression. This second constraint would force the finger apex to synchronize with the stressed vowel.

In this context, the observed pattern of synchronization can be interpreted as satisfying the required alignment of the verbal deictic site with the pointing gesture while maintaining a large range for hand-jaw coordination. Indeed, while coordination is rather strict for the first-stress condition, in order to remove any ambiguity in terms of verbal deictic site, it is in fact sufficient in the second-stress one that the apex arrives somewhere inside the second jaw opening gesture, even if it is not on the very apex. The coordination of the return point with the jaw apex would just indicate that the deictic task is now achieved, and the return movement can now be operated. In this scenario, it would be interesting to test what happens with three or more syllables, to see whether the coordination of apex with jaw apex is respected for all syllables except the last one, for which coordination would be relaxed into a coordination with the return gesture.

Acknowledgments

This work is part of the "Patipapa" project funded by the French Ministry of Research (Action Concertée Incitative "Systèmes Complexes en Sciences Humaines et Sociales"). We also wish to thank Christian Abry for his comments on this work.

4. References

Abry, C., Vilain A. and Jean-Luc Schwartz J.L. Introduction: Vocalize to Localize? A call for better crosstalk between auditory and visual communication systems researchers: From meerkats to humans. In Abry C., Vilain, A. and Schwartz, J.L., editors, *Vocalize to Localize*, 313–325, 2005.

Ducey-Kaufmann, V., Abry, C. & Vilain, C. When the Speech Frame meets the Sign Frame in a developmental framework. In *Proceedings of Emergence of Language Abilities: Ontogeny and phylogeny*, Lyon, 2005.

Feyereisen, P. The competition between gesture and speech production in dual-task paradigms. *Journal of Memory and Language*, 36(1):13-33, 1997.

Haviland, JB. Pointing, gesture spaces, and mental maps. In McNeill, D., editor, *Language and gesture,* 13–46, 2000.

Holender, D. Interference between a vocal and a manual response to the samestimuls. In Requin, J. & Stelmach, G.E., editors, *Tutorials in Motor Behavior*, 421-432. North Holland Publishing Company, 1980.

Levelt, W. J. M., Richardson, G. and La Heij, W. Pointing and voicing in deictic expressions. *Journal of Memory and Language*, 24:133-164, 1985.

MacNeilage, P.F. The frame/content theory of evolution of speech production. *Behav. Brain Sci.*, 21:499-511, 1998.

Nelson, W.L., Perkell, J.L. and Westbury, J.R. Mandibule movements during increasingly rapid articulations of single syllables : Preliminary observations. *J. Acoust. Soc. Am.*, 75(3):945-951, 1984.