

Pointing to a target while naming it with /pata/ or /tapa/: the effect of consonants and stress position on jaw-finger coordination

Amélie Rochet-Capellan¹, Jean-Luc Schwartz¹, Rafael Laboissière^{2,3}, Arturo Galván²

¹ GIPSA-Lab, Département parole et cognition (ICP, Institut de la Communication Parlée), France

² Max Planck Institute for Human Cognitive and Brain Science, Munich, Germany

³ Espace et Action, U864 INSERM, Lyon, France

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

Abstract

This study investigates jaw-finger coordination in a task consisting in pointing to a target while naming it with a /pata/ or a /tapa/ utterance stressed either on the first ('CVCV) or on the second (CV'CV) syllable. Optotrak measurements of jaw and finger displacements show that for 'CVCV names, the moment at which the finger reaches the target alignment is synchronized with the maximum of the first jaw opening motion. For CV'CV names, the synchronization occurs between the moment at which the finger leaves the target-alignment position and the maximum of the jaw opening motion for the second vowel. This pattern of synchronization does not depend on the target position or on the consonants order. These results add some support to theories involving the coordination of orofacial and brachiomanual gestures in the development and phylogeny of human languages. They call for more investigations on the link between speech and brachiomanual gestures in face-to-face communication.

Index Terms: speech-hand coordination, jaw-finger coordination, deixis, pointing, Optotrak.

1. Introduction

1.1. Substance-based deixis

The deictic function of language (refer to an object) has been assumed to be first supported by the hand and later connected to speech [1][2]. This connection could constitute a substrate for the emergence of language indexical signs (e.g. “this”) in the course of ontogeny and phylogeny. Hence, a close link might exist between speech and pointing deictic sites, enabling to *show* respectively by the hand and by the voice. Synchronization of speech and hand pointing in face-to-face communication [3][4] could consist in coherence between the speech focus and the pointing focus, with a trend to align the arm-hand-finger system with the speech focus. Moreover, developmental studies displayed a link between the frequencies of arm-hand gestures and jaw oscillations in babbling [5]. This link would evolve toward a “rendez-vous” between the motor control of jaw oscillations in babbling, the *speech frame* [6][7] and brachiomanual oscillations in pointing, the *sign frame* [8]. This “rendez-vous” would play a key role in language acquisition [9][10]. In addition, it leads to suspect that synchronization between speech and pointing gestures could rest on a close coordination of the jaw and arm-hand oscillatory systems, anchored in motor control acquisition.

1.2. Focus and jaw-finger coordination

In this framework, we previously studied jaw-finger coordination in a task consisting in pointing to a target while naming it with a CVCV word (/papa/ vs. /tata/) stressed either on the first ('CVCV) or on the second (CV'CV) syllable [11]. The hypothesis was that the instant of finger-target alignment (pointing apex, P_A) would be synchronized with the instant of maximum jaw opening (apex) for the stressed vowel; that is, with the first jaw apex (J_{A1}) for 'CVCV names and the second one (J_{A2}) for CV'CV ones. The results showed that P_A was actually close to J_{A1} for 'CVCV names. For CV'CV ones, P_A occurred after J_{A1} , but before J_{A2} , roughly at an equal delay between J_{A1} and J_{A2} , contrary to the prediction. However, in that case, J_{A2} was synchronized with the end of the period of finger-target alignment, when finger began its return stroke. Hence, there seems to be clear jaw-finger coordination, though different from our original hypothesis. The effect of stress position on finger-jaw synchronization mainly results from a speech adaptation: jaw movement begins earlier for CV'CV than for 'CVCV names. However, it also involves the duration during which the finger points on the target: the finger stays longer in its apex position for CV'CV than 'CVCV names, “expecting”, in some sense, jaw arrival on the stressed vowel. Further analyses showed that stress modifies amplitude and duration of the two jaw opening gestures, but has no effect on initiation and apex times of the pointing gesture. Conversely, increasing target distance delays both pointing apex and jaw apices, but has no effect on amplitude and duration of the two jaw opening strokes. Altogether, this previous experiment showed that the speech focus tends to be achieved during the finger-target alignment and that jaw events tend to be anchored in pointing events. Moreover, the pointing gesture aims at reaching the target alignment; it is mainly “target-driven”. On the contrary, jaw motions are adapted both to the phonetic goal and to the pointing task.

1.3. Jaw profiles and jaw-finger coordination

The previous study used CVCV words with a single consonantal articulation place. However, in another study focused on jaw-tongue-lips coordination in speech sequences, we found different jaw motion profiles in duplicated (/tata-/papa/) vs. variegated (/pata-/tapa/) utterances produced with rate increase [12][13]. The main tendency is to realize variegated CVCV sequences with large asymmetries between the two jaw cycles, evolving towards a single jaw cycle with rate increase. On the contrary, duplicated CVCV always required two jaw cycles. The present study aims at testing if the pattern of jaw-finger coordination observed with /papa/ and /tata/ with different stress positions resists to the

modifications in jaw dynamics imposed by variegation in /pata/ and /tapa/ sequences. This would provide an argument for the independence of internal speech planning relative to the adaptation of speech gestures to pointing gestures.

2. Methods

Participants. Twenty native Brazilian Portuguese speakers (4 men and 16 women), right-handed with normal or corrected-to-normal vision, participated in 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. The participants and the procedure were the same as those of our previous study [11].

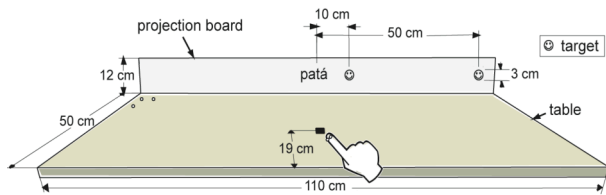


Figure 1. *Experimental setup*

Procedure. The participant was seated at a table. She/he was informed that a word and a red smiley ☺ sign (the target) would appear on the board in front of her, projected by a beamer. The word was introduced as the name of the person represented by the smiley-target. It could be either /pata/ or /tapa/ with a stress on the first (e.g. /pàta/) or on the second (e.g. /patà/) vowel. The vowel /a/ was selected since its realization requires a large jaw opening gesture. The target appeared in the right visual field, either at the near or at the far position (see Figure 1). The task was to point to the smiley and name it with the word as soon as its color changed from red to green. 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. A black square on the midline of the table indicated the onset position for the finger-pointing gesture. The experiment was divided into four blocks separated by a 30 s pause. A block contained 4 practice trials and 40 experimental trials, 10 for each [stress position]*[target position]*[consonants order] experimental condition. The order of the trials was randomized for each block and each participant.

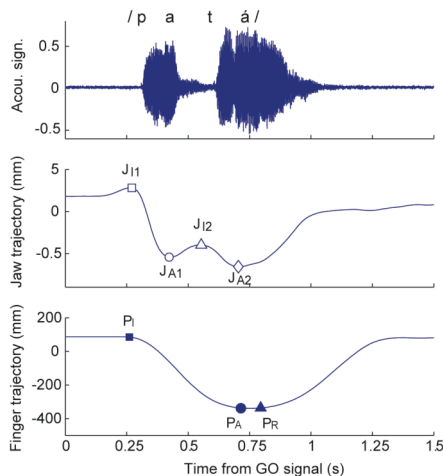


Figure 2. *Record and labeling for a /patà/ trial.*

Data recording and processing. Finger and jaw movements were recorded using an Optotrak. Two 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 markers were pasted on the table to provide a referential for the moving markers. Head motion was measured by three markers attached to a plastic triangle fixed around the subject's head. Jaw positions were computed in relation to the head moving reference frame. The markers positions were sampled against time at 100 Hz. In order to reduce the amount of data to process, we conducted a Principal Component Analysis (PCA) on the Optotrak data. The first principal component appears as a good representation of markers motions. In average, for the 20 participants, it explained 98% of the variance for the finger markers and 95% for the jaw marker. These signals were lowpass filtered at 15 Hz with a Butterworth filter. The temporal measurements are indicated in Figure 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.

3. Results

Absolute positions of jaw and finger events are displayed in Figure 3. As in [11], we also computed positions of P_A and P_R relative to the two jaw apices, defined as:

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

A value of 0 (respectively 1) indicates that the finger event is synchronized with J_{A1} (respectively J_{A2}). Finger positions relative to jaw apices are displayed in Figure 4. Amplitude and duration of jaw and finger strokes, were also computed as the difference of position and duration between the initiation and the apex of the stroke, respectively. They are displayed in Figure 5. The effects of the three experimental factors on each measure were tested using three-way within-subject ANOVAs (effects are considered significant for $p < .05$ with $F(1, 19)$).

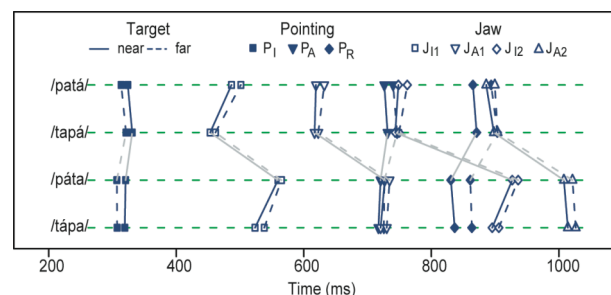


Figure 3. *Mean durations from the GO-signal to jaw and finger events according to experimental conditions.*

3.1. Pattern of jaw-finger synchronization

The effect of stress position on $P_{A/J}$ is significant, with $P_{A/J}$ smaller for 'CVCV (-0.02) than for CV'CV (0.41) ($p < .0001$). The effect of stress position on $P_{R/J}$ is also significant, with $P_{R/J}$ smaller for 'CVCV (0.44) than for CV'CV (0.95) ($p < .0001$). Target position and consonant order effects on $P_{A/J}$ are not significant, but $P_{R/J}$ depends on the target position

with greater value in the far (0.73) than in the near (0.66) target condition ($p < .01$). Interaction effects are not significant.

Hence, the synchronization pattern is similar to the one found previously [11]. Firstly, the pointing apex is close to the first jaw apex for 'CVCV sequences while it occurs at about an equal delay from the two jaw apices for CV'CV ones. Secondly, the pointing return occurs close to the second jaw apex for CV'CV names while it occurs at about an equal delay from the two jaw apices for 'CVCV ones. This pattern of synchronization does not depend on target position or consonant order.

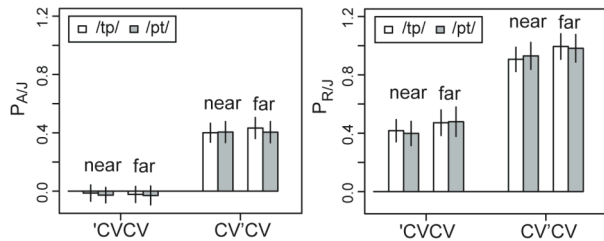


Figure 4. Position of pointing apex ($P_{A/J}$, left) and pointing return onset ($P_{R/J}$, right) relative to jaw apices according to experimental conditions.

3.2. Finger motion

Increase in target distance results in a significant 11 ms advance of P_I , and a significant delay of P_A and P_R (respectively 11 ms and 30 ms, $p < .05$). Thus, the pointing forward motion (from P_I to P_A) is 20 ms longer in the far- than in the near- target condition ($p < .0001$; Figure 5, top right). The stroke amplitude is also 147 mm greater in the far- than in the near- target condition ($p < .0001$, Figure 5, top left). Finally, the plateau ($P_R - P_A$) duration is 19 ms longer in the far- than in the near- target condition ($p < .01$, see Figure 3).

The stress position has no significant effect on P_I and P_A and on the amplitude and the duration of the forward motion. On the contrary, P_R is 35 ms later for CV'CV than for 'CVCV sequences ($p < .001$). Hence the plateau ($P_R - P_A$) duration is 22 ms longer for CV'CV than 'CVCV sequences ($p < .001$).

The consonant order has no significant effect on P_I and P_A and on amplitude and duration of the forward motion. Then, this factor has a significant but weak effect on the ($P_R - P_A$) duration (4 ms longer for /tp/ than /pt/, $p < .05$).

No interaction effect is significant.

Altogether, the finger forward motion just depends on the target position. Then, once the alignment with the target is achieved, the plateau duration adapts to the stress position: the finger can wait for the jaw.

3.3. Jaw motion

Figure 3 shows that all jaw events J_{I1} , J_{A1} , J_{I2} and J_{A2} occur significantly earlier for CV'CV than 'CVCV names ($p < .001$). The J_{A1} to J_{A2} duration is larger for 'CVCV utterances (290 ms) than for CV'CV ones (275 ms). Duration of the first jaw motion is 35 ms greater for 'CVCV than for CV'CV utterances ($p < .0001$). Conversely, the second jaw motion is 44 ms longer for CV'CV than for 'CVCV utterances ($p < .0001$). Similarly, the amplitude of the first stroke is 3.1 mm larger for 'CVCV than for CV'CV utterances ($p < .0001$) while the amplitude of the second stroke is 3.8 mm larger for CV'CV than for 'CVCV utterances ($p < .0001$). The consonant order has a significant effect on J_{I1} and J_{I2} , occurring respectively 34 ms ($p < .001$) and 20 ms ($p < .01$) earlier for /tp/ than for /pt/, but no significant effect on J_{A1} and J_{A2} . It does not significantly affect the first stroke amplitude while

the second stroke is 3.2 mm larger for /tp/ than /pt/ ($p < .0001$). Furthermore, the first and second jaw strokes are, respectively, 29 ms ($p < .001$) and 26 ms ($p < .0001$) longer for /tp/ than /pt/. The interaction between stress and consonant order is significant only for J_{I2} with a greater stress effect for /pt/ than for /tp/ ($p < .05$).

Jaw events occur about 10 ms earlier in the near- than in the far- target condition. This target effect is significant for J_{A1} , J_{I2} and J_{A2} ($p < .05$) but not for J_{I1} . Moreover, it does not significantly interact with the two other factors and it has no significant effect on amplitude and duration of jaw strokes.

Altogether, jaw motion is mainly driven by the phonetic goal with different profiles according to the stress position and to the consonant order. The timing of jaw events is also adapted to the finger pointing motion, as showed by the target position effects on jaw timing.

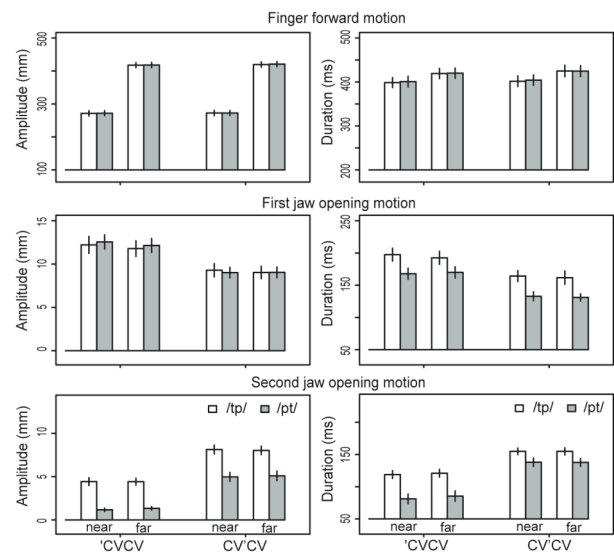


Figure 5. Amplitude and duration of finger forward stroke (top) and of first (middle) and second (bottom) jaw opening strokes according to experimental conditions.

4. Discussion

4.1. A stable jaw-finger coordinative pattern

Globally, the results confirm those we previously obtained with /papa/ and /tata/ target names [11]. First, the stress position has the same effect on jaw-finger synchronization. The jaw apex for the stressed vowel is included in the period of finger-target alignment. More precisely, for 'CVCV names, the jaw apex for the stressed vowel is aligned with the onset of finger-target alignment. For CV'CV names, the jaw apex for the stressed vowel is aligned with the end of finger-target alignment. This pattern of synchronization mainly results from both an adaptation of the jaw motion and of the duration for which the finger points to the target. Indeed, jaw response begins earlier for CV'CV as compared to 'CVCV. This lead makes the second jaw apex closer to the pointing apex. However, it is not enough to make the second jaw apex synchronized with the pointing apex. Consequently, the finger has to stay longer in its apex position for CV'CV names as compared to 'CVCV ones: it waits for speech focus achievement. This mainly results from the fact that the pointing motion for reaching the target alignment is "target driven": it may begin as soon as possible in order to reach the target. This "externally-driven" system would take priority

over the “internally-driven” and more flexible speech system, which agrees with previous results [4][14][15].

4.2. How does jaw motion adapt to the phonetic and deictic requirements?

The jaw motion is affected by the stress position, the consonant order and the target position. The stress position modifies both the motion timing and parameters. The initiation and apex of opening strokes occur earlier for CV'CV than for 'CVCV sequences. Unsurprisingly, the duration and amplitude of the first stroke are greater for 'CVCV than for CV'CV utterances while the reverse is observed for the second stroke. The consonant order affects the initiation times of jaw opening strokes but not their apices times. It also affects the amplitude and duration of jaw strokes with a tendency to observe longer opening strokes for /tp/ than for /pt/ and a larger second opening stroke for /tp/ as compared to /pt/. However these changes do not affect jaw-finger apices synchronization.

The target position mainly affects the timing of jaw motion. The lag of initiation and apex times of the two jaw strokes from the near- to the far- target condition is 10 ms, which is about that observed in [11] and corresponds to the delay of the finger apex from the near- to the far- target. Hence, the increase of target distance induces a translation in time of jaw motion while amplitudes and durations remain the same. This adaptation allows preserving the jaw-finger synchronization pattern for both target positions.

5. Conclusion

Taken together, these results agree with the idea of an anchoring of speech focus in pointing gesture in deictic expressions. Moreover, this anchoring seems to be supported by a synchronization of the *speech-frame* with the *sign-frame* [8]. This happens to result from two independent levels of speech adaptation. Firstly, the effects of stress position and consonant order indicate an “internal” adaptation of the gestures to the phonetic goal. Secondly, the spatial target effect suggests an “external” adaptation for the synchronization with the pointing motion. These results agree with those obtained on /papa/ and /tata/ sequences [11].

Within the hotly debated topic of the gestural vs. vocal origin of human language [16], recent theories suggest a reconciliation of gestures and vocalisations in proposals where the coordination of sounds and gestures would be a crucial step [1][17]. Future experiments allowing to better understand the coordination of the orofacial and brachiomanual systems in face-to-face communication should provide very important inputs in this framework.

6. Acknowledgements

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.

7. References

[1] 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.

- [2] Haviland, JB. Pointing, gesture spaces, and mental maps. In McNeill, D., editor, *Language and gesture*, 13–46, 2000.
- [3] McNeill, D. Action, thought and language. *Cognition* 10: 201-208, 1981.
- [4] 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.
- [5] Petitto L.A., Holowka S, Sergio L.E. and Ostry D.J. Language rhythms in baby hand movements. *Nature*, 413: 35-6, 2001.
- [6] MacNeilage, P.F. and Davis B.L. On the origins of internal structure of word forms. *Science*, 288:527-531, 2000.
- [7] MacNeilage, P.F. The frame/content theory of evolution of speech production. *Behav. Brain Sci.*, 21:499-511, 1998.
- [8] 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.
- [9] Pizzuto, E., Capobianco, M., & Devescovi, A. Gestural-vocal deixis and representational skills in early language development. In C. Abry, A. Vilain & J.-L. Schwartz (Eds.) *Special issue: "Vocalize to Localize II". Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, 6 (2), 223-252, 2005.
- [10] Volterra, V. , Caselli, M. C. , Capirci, O. , Pizzuto, E. Gesture and the emergence and development of language. In M. Tomasello and D. Slobin, (Eds.) *Beyond nature-nurture – Essays in honor of Elizabeth Bates*. Mahwah, N. J. : Lawrence Erlbaum Associates, pp. 3-40, 2005.
- [11] Rochet-Capellan, A., Schwartz J-L., Laboissière, R. and Galván, A., Finger-jaw coordination during a deictic gesture with CVCV utterances: the effect of stress position, in *Proceedings of ISSP2006*, 193-200, 2006.
- [12] Rochet-Capellan, A. and Schwartz, J.-L. The Labial-Coronal effect and CVCV stability during reiterant speech production: An articulatory analysis, in *Proceedings of INTERSPEECH-2005*, 1013-1016, 2005.
- [13] Rochet-Capellan, A. and Schwartz, J.-L. (under revision). An articulatory basis for the Labial-to-Coronal effect: /pata/ seems a more stable articulatory pattern than /tapa/.
- [14] Holender, D. Interference between a vocal and a manual response to the same stimuli. In Requin, J. & Stelmach, G.E., editors, *Tutorials in Motor Behavior*, 421-432. North Holland Publishing Company, 1980.
- [15] Feyerisen, P. The competition between gesture and speech production in dual-task paradigms. *Journal of Memory and Language*, 36(1):13-33, 1997.
- [16] Corballis, M.C. From mouth to hand : Gesture, speech and the evolution of right-handedness. *Behavioral and Brain Sciences*, 26, 199-260, 2003.
- [17] Arbib, M. A. Interweaving protosign and protospeech: Further developments beyond mirror. In C. Abry, A. Vilain & J.-L. Schwartz (Eds.) *Special issue: "Vocalize to Localize II". Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, 6 (2), 145-171, 2005.