

Why does picture naming take longer than word reading? The contribution of articulatory processes

Stéphanie Riès · Thierry Legou · Borís Burle ·
F.-Xavier Alario · Nicole Malfait

Published online: 30 June 2012
© Psychonomic Society, Inc. 2012

Abstract Since the 19th century, it has been known that response latencies are longer for naming pictures than for reading words aloud. While several interpretations have been proposed, a common general assumption is that this difference stems from cognitive word-selection processes and not from articulatory processes. Here we show that, contrary to this widely accepted view, articulatory processes are also affected by the task performed. To demonstrate this, we used a procedure that to our knowledge had never been used in research on language processing: response-latency fractionating. Along with vocal onsets, we recorded the electromyographic (EMG) activity of facial muscles while participants named pictures or read words aloud. On the basis of these measures, we were able to fractionate the verbal response latencies into

two types of time intervals: premotor times (from stimulus presentation to EMG onset), mostly reflecting cognitive processes, and motor times (from EMG onset to vocal onset), related to motor execution processes. We showed that premotor *and* motor times are both longer in picture naming than in reading, although articulation is already initiated in the latter measure. Future studies based on this new approach should bring valuable clues for a better understanding of the relation between the cognitive and motor processes involved in speech production.

Keywords Speech production · Naming · Reading · Verbal response times · Articulation · Electromyography · Psycholinguistics · Motor control

Electronic supplementary material The online version of this article (doi:10.3758/s13423-012-0287-x) contains supplementary material, which is available to authorized users.

S. Riès
Laboratoire de Neurosciences Cognitives, UMR 7291 &
Laboratoire de Psychologie Cognitive, UMR 7290,
CNRS / Aix-Marseille Université,
3, place Victor-Hugo,
13331 Marseille cedex 3, France

S. Riès
Helen Wills Neuroscience Institute & Department
of Psychology,
Knight laboratory University of California 210 Barker Hall,
Berkeley, CA 94720-3190, USA
e-mail: stephanie.riès@berkeley.edu

T. Legou
Laboratoire Parole et Langage, UMR 7309,
CNRS / Aix-Marseille Université,
5, avenue Pasteur,
13100 Aix-en-Provence, France
e-mail: thierry.legou@lpl-aix.fr

B. Burle
Laboratoire de Neurosciences Cognitives, UMR 7291,
CNRS / Aix-Marseille Université,
3, place Victor-Hugo,
13331 Marseille cedex 3, France
e-mail: boris.burle@univ-amu.fr

F.-X. Alario
Laboratoire de Psychologie Cognitive, UMR 7290,
CNRS / Aix-Marseille Université,
3, place Victor-Hugo,
13331 Marseille cedex 3, France
e-mail: francois-xavier.alario@univ-amu.fr

N. Malfait (✉)
Institut de Neurosciences de la Timone, UMR 7289,
CNRS / Aix-Marseille Université,
27, boulevard Jean Moulin,
13385 Marseille cedex 5, France
e-mail: nicole.malfait@univ-amu.fr

Response latencies are longer for naming pictures than for reading words aloud. This established fact has been a focus of interest in psycholinguistics for over a century (Cattell, 1885; Ferrand, 1999). It has been interpreted in different ways. Some studies have favored the so-called “semantic hypothesis,” according to which time-consuming access to semantic information is only required in picture naming (Theios & Amrhein, 1989). According to the “uncertainty hypothesis,” the effect is due to a more equivocal stimulus–response association in picture naming than in word reading. A picture, but not a printed word, can be described with different labels (Ferrand, 1999; Fraisse, 1969). Underlying these hypotheses is the idea that performance differences (e.g., in verbal response times) are entirely due to central cognitive stages, such as differences in word-selection processes, while articulatory processes would be insensitive to the nature of the task (e.g., Levelt, Roelofs, & Meyer, 1999; Roelofs, 2004).

Kello, Plaut, and MacWhinney (2000) and Damian (2003) investigated the relationship between the cognitive and motor processes involved in speech production using a Stroop task (Stroop, 1935). In addition to verbal response onsets, they measured verbal response durations. When high time pressure was imposed, a Stroop effect was observed on verbal response durations. According to Kello et al., this demonstrates that task demands can flexibly change speech production processes from “serial,” in which each processing stage must be over before the next one starts, to “cascaded,” in which articulatory processes can be initiated before semantic retrieval and phonological access are complete. In the latter case, how cognitive access occurs may influence the way that articulation is performed. Hence, this account questioned the dissociation between cognitive and motor processes that is generally taken for granted. However, Damian failed to replicate the critical effect reported by Kello et al., and argued on the contrary that in “conceptually driven” tasks, cognitive processes cannot cascade down to the articulatory stage.

Close consideration of word-form encoding processes suggests that words and pictures may differ in how they trigger a verbal response. In word reading, but not in picture naming (“conceptually driven” responses), partial sublexical information is allowed to permeate through to the output before preproduction processes are complete. Consistent with this idea, Hennessey and Kirsner (1999) showed that articulatory *durations* of the same word were *longer* for printed words than for pictures. In their view, this difference (observed for low-frequency items only) indicates that response execution can be stretched out online to compensate for an early start on the basis of partial information. In contrast to this view, however, some evidence supports the idea that reading a word aloud starts only when the computation of the phonology of the entire verbal response is

complete, at least with monosyllabic words (Rastle, Harrington, Coltheart, & Palethorpe, 2000).

Despite the apparent inconsistencies, these studies are valuable, as they address an issue that is seldom investigated in language-processing research—namely, the relationship between the cognitive and motor processes involved in speech production. We submit, however, that the measures most commonly used may have been suboptimal. On the one hand, verbal response times (RTs) reflect the sum of the processes occurring from stimulus presentation to vocal onset, without distinguishing between cognitive and motor processes. Indeed, verbal RTs include not only the time allocated to word selection, but also the time needed to plan and initiate the articulatory motor processes preceding vocal onset. On the other hand, verbal response durations are measured from vocal onset, and thus do not include the early cognitive phase.

In the frame of chronometric-analysis studies using manual responses, RT fractionating procedures have been developed, intended to help disentangle time delays that reflect response selection from delays related to motor response execution (e.g., Botwinick & Thompson, 1966; Burle, Possamaï, Vidal, Bonnet, & Hasbroucq, 2002). The idea is to break the RTs into two time intervals on the basis of the electromyographic (EMG) activity that can be recorded preceding the overt manual response (e.g., from the flexor pollicis brevis muscle for thumb buttonpress responses). The first delay, between the presentation of the stimulus and the onset of EMG activity, would essentially reflect stimulus encoding, response selection, and motor planning, and has been referred to as “premotor time” (pre-MT). The second time interval, which separates EMG onset from buttonpress, would characterize motor response execution processes and has been referred to as “motor time” (MT). Applying this procedure, these studies could demonstrate that task manipulations (e.g., the number of possible responses), initially thought to influence premotor processes only, also affect MTs (Possamaï, Burle, Osman, & Hasbroucq, 2002).

To our knowledge, RT-fractionating procedures have never been applied in the field of language production research (but see the discussion of a “not . . . observable” execution-acoustic interval in Rastle, Croot, Harrington, & Coltheart, 2005). Here we exploited this method as a means to test whether the motor response execution processes engaged in articulation are affected by task demands. Along with vocal onsets, we measured the electromyographic (EMG) activity of lip muscles as participants named pictures or read words aloud. On the basis of the latter measure, we divided verbal RTs into two time intervals: pre-MTs and MTs (Fig. 1). There is, however, a clear difference between the way that this method was originally applied to simple manual responses and the way that we exploited it here. In typical thumb buttonpress responses, there is a direct causal

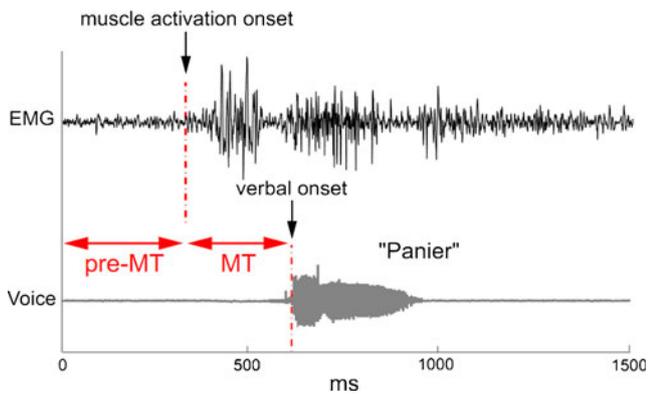


Fig. 1 Verbal response time fractionating: The standard naming latency is divided into premotor time (pre-MT, between the stimulus and electromyographic [EMG] onsets) and motor time (MT, between the EMG and vocal onsets)

link between flexor pollicis brevis EMG and thumb flexion. In contrast, EMG activities recorded from facial muscles and voice onsets correspond to different effectors. (Therefore, trials in which the EMG activities follow the vocal onset can also be observed; for details on this procedure on this procedure, see the [supplementary materials](#).) In addition, EMG activities trigger the action of effector muscles, whereas vocal onset is the effect of an effector. Thus, fractionating verbal RTs allowed us to assess whether task demands induce systematic variations in the time (phase) relationship between the actions of two vocal-tract effectors (protrusion and/or opening of the lips and glottal aperture).

Under the hypothesis that articulatory motor execution is insensitive to the origin of the speech to be produced, MTs should not be affected by the task performed. If, on the contrary, MTs differ across reading and naming, this finding would directly challenge the widely held idea. We tested the hypothesis in two similar experiments, both contrasting reading and naming, which differed as follows: In the first experiment, the two tasks were performed in different blocks involving different but carefully matched materials, whereas in the second experiment the words to be read corresponded to the names of the pictures, and the two types of trials were intermixed within the testing blocks. Other differences are detailed below.

Experiment 1

Method

Participants Ten native French speakers with normal or corrected-to-normal vision participated in the experiment (age: $M = 21$ years, $\sigma = 3$ years). The data from one participant were removed from the analysis due to voice-key problems.

Materials and design The stimuli were 40 line drawings of common objects selected from published collections (e.g., Alario & Ferrand, 1999) or designed for the experiment (name agreement: $M = 92\%$, $\sigma = 12\%$) and 40 French written words. The stimuli were presented on a CRT screen positioned 150 cm from the participant. The pictures were presented in black on an 11×11 cm white square, and the words were written in white on a black background, covering on average a comparable visual angle of 4° . The picture names' and the words' initial phonological properties (a first /b/, /f/, /m/, /p/, or /v/ phoneme, followed by an open palatal, closed palatal, or closed velar phoneme), lexical frequency, and number of syllables were matched (see the [supplementary materials](#)).

Each experimental run comprised 40 trials, which were either the 40 pictures or the 40 written words presented pseudorandomly. Each participant was tested on eight runs. The order of the runs was counterbalanced across participants.

Procedure Each trial consisted of the following sequence: (1) a fixation point (“+” sign) for 500 ms, (2) the picture or word presented until the participant responded or until a 1,500-ms deadline was reached, and (3) a blank screen for 2,000 ms, all controlled by E-Prime 2.0 (Psychology Software Tools, Inc., Sharpsburg, PA). In a preliminary familiarization phase, the 40 words and 40 drawings that would be used in the test phase were each presented once in two separate runs, with counterbalanced orders across participants. The instruction was to read or name each of the items. For the pictures, the participants were corrected when they produced an unexpected name. The actual test phase lasted about 30 min. Participants were asked to name the pictures and to read the words aloud as quickly and as accurately as possible.

Vocal responses were recorded with a piezoelectric microphone, sampled at 22050 Hz. EMG activity was recorded monopolarly with preamplified surface Ag/AgCl electrodes (BioSemi, Inc.) from three oro-facial muscles: *orbicularis oris*, *risorius*, and *mentalis*; the sampling rate was 2048 Hz (filters: DC to 268 Hz, 3 dB/octave). A passive reference electrode was placed on the earlobe, and two active reference electrodes were positioned on the wrist.

Data processing The EMG data were filtered offline (high pass = 10 Hz). Because of a poor signal-to-noise ratio, the EMG recorded from the *mentalis* muscle was discarded from the analyses. For offline detection of the vocal and EMG onsets, the same algorithm was used, one based on the Teager–Kaiser operator, which permits the detection of abrupt onsets of energy in the signal (Li, Zhou, & Aruin, 2007). Processed by this operator, the signal was then filtered (moving average window) to discard detections of

unexpected isolated pulses (e.g., due to lip opening). For each trial, the detected onset was then visually checked and readjusted when needed. This was performed blind to the task (reading or naming). Trials on which the baseline of the EMG channels was too noisy, as well as those on which the recording failed (e.g., because of faulty voice-key triggering or no verbal response within 1,500 ms) were excluded from further analysis. Trials were coded as errors if they included partial or complete production of incorrect or unexpected words, or included verbal disfluencies (stuttering, utterance repairs, etc.).

Mixed-effects analyses of variance (ANOVAs) were run on the three different measures (verbal RTs, pre-MTs, and MTs) with the factors Task (naming vs. reading), Muscle (*orbicularis* or *risorius*), and First Phoneme (/b/, /f/, /m/, /p/, or /v/) as fixed effects, and with participants (F_1) or linguistic items (F_2) as random variables. For the linguistic-item analysis, Task and First Phoneme were between-items factors, and muscle was a within-item variable. Student's two-tailed t tests were used for post-hoc pairwise comparisons. For all tests, a significance level of .05 was used. The p values were adjusted for multiple comparisons using the Bonferroni correction.

Results On average, 4.5 % of the trials ($\sigma = 3.1$ %) per participant were rejected due to poor EMG signal-to-noise ratios. The results are summarized in Fig. 2. Verbal RTs were longer for pictures than for words [$\Delta = 185$ ms, $\sigma = 49$ ms; $F_1(1, 8) = 139.16, p < .001, F_2(1, 70) = 484.15, p < .001$]. The error rate was higher for pictures (3.56 %, $\sigma = 3.73$ %) than for words (0.65 %, $\sigma = 0.98$ %) [$t_1(8) = 2.90, p < .05; t_2(52.45) = 3.49, p < .01$]. We found a significant main effect of First Phoneme [$F_1(4, 32) = 4.66, p < .01; F_2(4, 70) = 4.31, p < .01$], due to utterances starting with /p/ being associated with longer RTs, although none of the pairwise comparisons reached significance [/p/ vs. /v/, $t_1(8) = 2.07, p = .29$; all other comparisons, $t_1s(8) < 1, t_2s < 1$].¹ The interaction between task and First Phoneme was not significant [$F_1(4, 32) = 1.24, p = .31; F_2(4, 70) < 1$].

Pre-MTs were longer for pictures than for words [$\Delta = 164$ ms, $\sigma = 45$ ms; $F_1(1, 8) = 115.88, p < .001; F_2(1, 70) = 549.82, p < .001$]. No effect of muscle emerged [$F_1(1, 8) < 1; F_2(1, 70) = 2.76, p = .10$], and no interaction between Task and Muscle [$F_1(1, 8) = 1.45, p = .26; F_2(1, 70) = 2.23, p = .14$]. A main effect of First Phoneme [$F_1(4, 32) = 3.18, p < .05; F_2(4, 70) = 4.94, p < .01$] was observed. However, while /f/ and /v/ tended to produce longer responses, the pairwise differences did not reach significance [all $t_1s(8) < 1.21, t_2s < 1$].¹ We found no interactions between First Phoneme and task [$F_1(4, 32) = 1.10, p = .37; F_2(4, 70) < 1$] or

between First Phoneme and muscle [$F_1(4, 32) = 1.82, p = .15$; however, this interaction was significant by items: $F_2(4, 70) = 5.83, p < .001$], and no three-way interaction [$F_1(4, 32) < 1; F_2(4, 70) = 1.36, p = .26$].

Critically, MTs were longer in naming pictures than in reading words [$\Delta = 23$ ms, $\sigma = 30$ ms; $F_1(1, 8) = 5.10, p = .05; F_2(1, 70) = 26.58, p < .001$]. We found no main effect of muscle² [$F_1(1, 8) < 1; F_2(1, 70) = 2.76, p = .10$] and no interaction between muscle and task [$F_1(1, 8) = 1.45, p = .26; F_2(1, 70) = 2.23, p = .14$], but a significant main effect of First Phoneme did emerge [$F_1(4, 32) = 11.46, p < .001; F_2(4, 70) = 27.71, p < .001$]. Longer MTs were observed for utterances starting with /p/ [p/ vs. /b/, $t_1(8) = 5.75, p < .01; t_2(29.85) = 6.79, p < .001$; other $t_1s(8) < 1, t_2s < 1.30$].¹ There were no interaction between First Phoneme and task [$F_1(4, 32) < 1; F_2(4, 70) < 1$] or First Phoneme and muscle [$F_1(4, 32) = 1.82, p = .15$; but significant by items: $F_2(4, 70) = 5.83, p < .001$], and no three-way interaction [$F_1(4, 32) < 1; F_2(4, 70) = 1.36, p = .26$]. Further details about interparticipant and intertrial variability are provided in the [supplementary materials](#).

Experiment 2

Method

Participants A group of 18 native French speakers with normal or corrected-to-normal vision participated in the experiment (age: $M = 20.6$ years, $\sigma = 1.5$ years). The data of six participants were excluded from the analysis due to over-noisy EMG recordings.

Materials and design The stimuli were 12 line drawings of common objects (name agreement: $M = 96$ %, $\sigma = 6$ %) and 12 French written words corresponding to the names of the pictures (see the [supplementary materials](#)). The First Phonemes were /b/, /m/, and /p/. Items beginning with /f/ or /v/ were not used, given the variability that they had introduced in Experiment 1. The type of second phoneme (open palatal, closed palatal, or closed velar) was controlled as in the previous experiment.

The familiarization phase (see Exp. 1) was followed by three testing runs in which all 12 pictures and 12 words were

¹ The degrees of freedom for the by-items t tests varied from 22.99 to 29.96.

² The estimates for the effects involving the Muscle factor were not expected to differ between the pre-MT and MT analyses. This is because the performance for all of the different muscles was estimated within trials, with the same RTs for all muscles. In other words, pre-MT + MT was constant across muscles. As a consequence, the variance associated with this factor was the same for pre-MTs and for MTs. Importantly, this was not the case for the factors Task and First Phoneme, for which the estimates involved different trials.

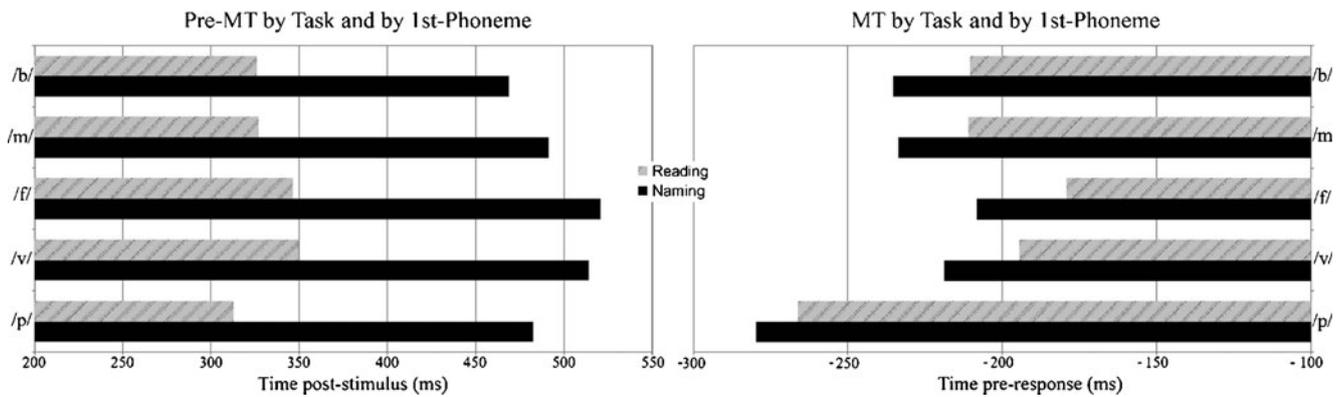


Fig. 2 For Experiment 1, premotor (pre-MT) and motor (MT) times across First Phonemes in the reading and naming tasks, averaged over muscles. As expected, pre-MTs are longer for naming than for reading

presented four times each (96 trials in total), intermixed in a pseudorandom order (stimuli involving the same response were separated by at least two other items).

Procedure The procedure was the same as in Experiment 1, except for the following modifications. The duration of the fixation point varied randomly between 500 and 1,000 ms. Voice and EMG were recorded by the same device (Keithley Instruments, Inc.). Bipolar montages of 6-mm-diameter Ag/AgCl surface electrodes (Grass Technologies, Inc.) were used to record EMG activity from four facial muscles: *levator labii superioris*, *risorius*, *orbicularis oris*, and *depressor labii inferioris*. (Recording of the *mentalis* muscle activity was discarded, as only a poor signal-to-noise ratio had been obtained in Exp. 1.) The sampling rate was 2000 Hz, and the reference electrode was placed over the left collarbone. The acoustic signal was recorded at 28000 Hz. Stimulus presentation durations were kept constant (1,500 ms) across trials, and both pictures and words were presented in black on a white background.

Data processing The acoustic and EMG data were analyzed as in Experiment 1, except that, for the linguistic-item analysis, First Phoneme was a between-items factor, while both Task and Muscle were within-item variables.

Results An average of 4.3 % of trials ($\sigma = 4.3$ %) were rejected due to poor EMG signal-to-noise ratios. Figure 3 summarizes the results. Again, verbal RTs were longer for pictures than for words [$\Delta = 90$ ms, $\sigma = 38$ ms; $F_1(1, 11) = 69.19$, $p < .001$; $F_2(1, 9) = 208.86$, $p < .001$]. Error rates were very low for both pictures and words [$M_s = 0.44$ % and 0.18 %, $\sigma_s = 1.32$ % and 0.33 %, respectively; pictures vs. words, $t_1(11) < 1$; $t_2(11) = -1.49$, $p = .17$]. There was an effect of First Phoneme [$F_1(2, 22) = 221.93$, $p < .001$; $F_2(2, 9) = 37.86$, $p < .001$] due to items starting with /p/ [p/ vs. /m/, $t_1(11) = 17.91$, $p < .001$; $t_2(11) = 7.29$, $p < .05$; /m/ vs. /

(left). Importantly, however, MTs are also longer for naming than for reading (right panel; note that MTs are depicted time-locked to speech onset)

b/, $t_1(11) < 1$; $t_2(11) < 1$]. The interaction between task and First Phoneme was not significant [$F_1(2, 22) = 1.43$, $p = .26$; $F_2(2, 9) < 1$].

Pre-MTs were longer for pictures than for words [$\Delta = 74$ ms, $\sigma = 29$ ms; $F_1(1, 11) = 77.46$, $p < .001$; $F_2(1, 9) = 143.84$, $p < .001$]. A main effect of muscle did emerge [$F_1(3, 33) = 15.66$, $p < .001$; $F_2(3, 27) = 314.43$, $p < .001$]: Pre-MTs were longer for *risorius* than for *orbicularis oris* [$t_1(11) = 3.99$, $p < .01$; $t_2(11) = 15.30$, $p < .001$; *orbicularis oris* vs. *depressor labii inferioris*, $t_1(11) = 2.28$, $p = .13$; $t_2(11) = 10.55$, $p < .001$; *levator labii superioris*, $t_1(11) < 1$; $t_2(11) = 3.33$, $p < .05$]. However, there was no main effect of First Phoneme [$F_1(2, 22) < 1$; $F_2(2, 9) < 1$]. A significant interaction was observed between Task and Muscle, with a larger effect of task for muscles for which the pre-MTs were larger [$F_1(3, 33) = 3.77$, $p < .05$; $F_2(3, 27) = 5.73$, $p < .01$; pairwise comparisons for *orbicularis oris* vs. *depressor labii inferioris*, $t_1(11) = 3.49$, $p < .05$; $t_2(11) = 3.47$, $p < .05$; other $t_1s(11) < 1$, $t_2s(11) < 1$]. The interaction between task and First Phoneme was marginal by participants [$F_1(2, 22) = 2.92$, $p = .07$; $F_2(2, 9) < 1$; all pairwise comparisons, $t_1s(11) < 1$, $t_2s(11) < 1.25$]. There was no interaction between First Phoneme and muscle [$F_1(6, 66) < 1$; $F_2(6, 27) < 1$] and no three-way interaction [$F_1(6, 66) = 1.69$, $p = .14$; $F_2(6, 27) = 1.61$, $p = .18$].

Confirming the results reported in Experiment 1, MTs were longer for pictures than for words [$\Delta = 16$ ms, $\sigma = 22$ ms; $F_1(1, 11) = 6.38$, $p < .05$; $F_2(1, 9) = 17.40$, $p < .01$]. We did find a main effect of muscle [$F_1(3, 33) = 15.67$, $p < .001$; $F_2(3, 27) = 316.10$, $p < .001$]: MTs were longer for *orbicularis oris* than for *risorius* [$t_1(11) = 3.99$, $p < .01$; $t_2(11) = 15.29$, $p < .001$; *depressor labii inferioris* vs. *orbicularis oris*, $t_1(11) = 2.28$, $p = .13$; $t_2(11) = 10.49$, $p < .001$; *risorius* vs. *levator labii superioris*, $t_1(11) < 1$; $t_2(11) = 3.32$, $p < .01$]. A significant interaction between Task and Muscle indicated that the effect of task was larger for muscles for which MTs were larger [$F_1(3, 33) = 3.78$, $p < .05$; $F_2(3, 27) = 5.72$, $p < .01$];

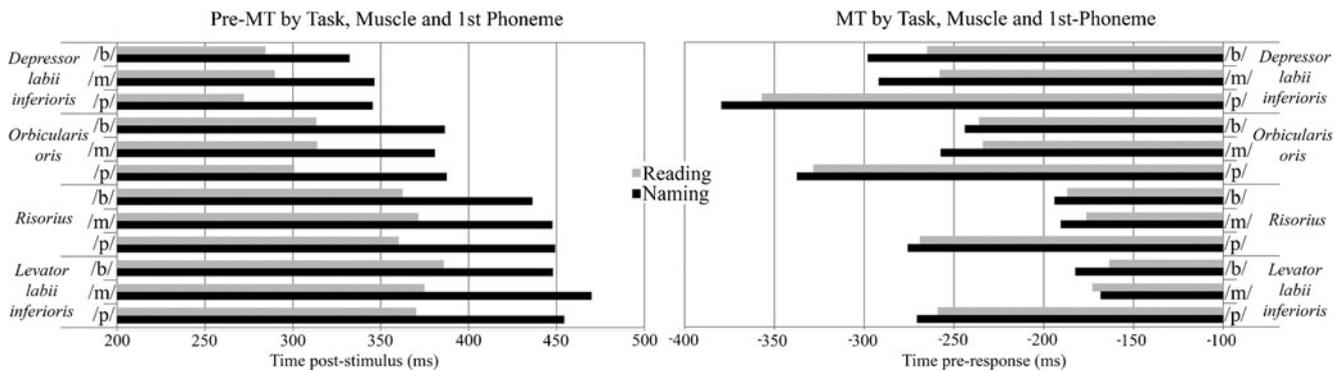


Fig. 3 Premotor (pre-MT) and motor (MT) times, broken down by task, muscle, and First Phoneme in Experiment 2. Here, too, both pre-MTs and MTs are affected by the task

pairwise comparison of *depressor labii inferioris* vs. *orbicularis oris*, $t_1(11) = 3.45$, $p < .05$; $t_2(11) = 3.45$, $p < .01$; other $t_1s(11) < 1$; $t_2s(11) < 1$. MTs were also affected by First Phoneme [$F_1(2, 22) = 127.69$, $p < .001$; $F_2(2, 9) = 31.03$, $p < .001$], as items starting with /p/ were slower [/p/ vs. /b/, $t_1(11) = 17.04$, $p < .001$; $t_2(11) = 9.94$, $p < .01$; /b/ vs. /m/, $t_1(11) < 1$; $t_2(11) < 1$]. The effect of First Phoneme did not interact with task [$F_1(2, 22) < 1$; $F_2(2, 9) < 1$] nor with muscle [$F_1(6, 66) < 1$; $F_2(6, 27) < 1$]. There was no three-way interaction [$F_1(6, 66) = 1.69$, $p = .14$; $F_2(6, 27) = 1.60$, $p = .18$].

Discussion

Our main finding was that the well-known difference between naming and reading in verbal RTs is not due solely to increased pre-MTs, but also to a significant lengthening of MTs for naming as compared to reading. Contrary to a widely held hypothesis, motor processes involved in the execution of verbal responses are also affected by the task being performed. Dividing verbal RTs on the basis of EMG activity recorded from facial muscles allowed us to demonstrate that the nature of the stimulus can have an impact on the coordination of the articulatory “gestures” of the different effectors (lip and glottis; Browman & Goldstein, 1992).

One may argue that these results simply reflect differences in strategies. In Experiment 1, where word and picture naming were presented in separate blocks, participants may have set a priori different response thresholds for the two tasks. However, the same task effect was observed in Experiment 2, where words and pictures were intermixed. Thus, we suggest that an interpretation in which central, nonmotor effects spread out into the motor articulatory processes is more likely, in agreement with a cascaded model of information processing (McClelland, 1979). In this view, information is continuously accumulated through different processing stages, and a response is produced when the last stage has reached a pre-defined level of information accumulation. The critical point

is that, as soon as the first level starts accumulating information, it transmits downstream its partial output. If one level is affected by an experimental manipulation, this will also impact downstream levels. In the present context, a task-processing difference at a cognitive level will percolate down to affect motor processes.

Such a cascading hypothesis implies the concurrent engagement of cognitive processes and motor output, and a dynamic flow of information between the two. (For example, segments produced during errors show articulatory features that are biased toward the nonproduced target segment; thus, information about the target segment or close alternatives is thought to have cascaded to articulatory processes: Goldrick & Blumstein, 2006; McMillan & Corley, 2010; see also Kello et al. 2000.) The available evidence comprises two features that allow us to better specify some aspects of these dynamics. First, in our data, the task effect was larger, the earlier that a muscle was activated. If the task effect stems from cognitive processes that start before the onset of articulation, there would be a shorter delay between their initiation and the EMG onset for the muscles activated first. The cascading of the task effect would therefore be more perceptible (and thus, possibly larger) for those early muscles than for muscles activated later.

Second, there is a contrast between the direction of the task effects reported here (shorter MTs for words than for pictures) and those reported by Hennessey and Kirsner (1999; for low-frequency items, durations were shorter for pictures than for words). In their account, word durations are longer because articulatory execution is stretched out to accommodate the fact that it was triggered on the basis of partial information (e.g., word beginnings) that needs to be completed online. To provide a unitary account of both findings, one may speculate post-hoc that the articulation of a smaller chunk (e.g., a word onset in reading) will be swifter and more efficient than the articulation of a complete phonetic program (e.g., a whole word in picture naming), but that subsequent execution will be stretched to compensate for a speedy start. This interpretation raises the

intriguing possibility that online adjustments of articulation processes do occur and are measurable, even in tasks as simple as single-word production (see also Kawamoto, Liu, Mura, & Sanchez, 2008). We note, however, that the “sublexical” hypothesis has been partially challenged on different grounds (e.g., Rastle et al., 2000).

As noted earlier, influential current explanations for the naming–reading difference in verbal RTs are diverse, but all are tied to processes occurring before the initiation of articulation: semantic access (Theios & Amrhein, 1989), response selection as indexed by stimulus–response compatibility (Ferrand, 1999; Fraise, 1969), or partial phonological or phonetic access (i.e., the “sublexical hypothesis”; see the previous paragraph). Adjudicating between the “semantic,” “uncertainty,” and “sublexical” hypotheses on the basis of the present data is not straightforward. It is worth noting, however, that the uncertainty hypothesis (Fraise, 1969) finds support in some converging independent nonlinguistic evidence, as response uncertainty has been shown to affect MTs in manual response tasks (Possamaï et al., 2002).

In conclusion, we have shown that the well-known difference in verbal-RT performance between naming and reading is not due to the lengthening of cognitive processes solely, as has often been taken for granted. Our findings demonstrate that motor articulatory processes are also affected by speech production conditions, so the duration of the articulatory stage contributes in its own right to the age-old difference between latencies in word reading and picture naming. In addition, our study presents a procedure entirely new to research on language processing—verbal RT fractionating—that permits finer-grained analyses than are commonly performed. Future studies based on this new approach that is intended to capture finer-grained effects should bring valuable clues for a better understanding of the relation between cognitive and motor processes.

Author Note This research was supported by Grant No. ANR-07-JCJC-0074 from the Agence Nationale de la Recherche (France); by the European Research Council under the European Community’s Seventh Framework Program (FP7/2007-2013, Grant Agreement Nos. 241077 & 263575); by an MNRT Grant from the Ministère de l’Enseignement et de la Recherche (France) awarded to S.R.; and by IFR 131, Sciences du Cerveau et de la Cognition (France). We thank Licia Damiani for collecting the data.

References

- Alario, F.-X., & Ferrand, L. (1999). A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments, & Computers*, *31*, 531–552. doi:10.3758/BF03200732
- Botwinick, J., & Thompson, L. W. (1966). Premotor and motor components of reaction time. *Journal of Experimental Psychology*, *71*, 9–15.
- Browman, C., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica: International Journal of Speech Science*, *49*, 155–180.
- Burle, B., Possamaï, C. A., Vidal, F., Bonnet, M., & Hasbroucq, T. (2002). Executive control in the Simon effect: An electromyographic and distributional analysis. *Psychological Research*, *66*, 324–336.
- Cattell, J. M. (1885). Über die zeit der Erkennung und Benennung von Schriftzeichen, Bildern und Farben [The time it takes to recognize and name letters, pictures, and colors]. *Philosophische Studien*, *2*, 635–650.
- Damian, M. F. (2003). Articulatory duration in single word speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 416–431.
- Ferrand, L. (1999). Why naming takes longer than reading? The special case of Arabic numbers. *Acta Psychologica*, *100*, 253–266.
- Fraise, P. (1969). Why naming is longer than reading? *Acta Psychologica*, *30*, 96–103.
- Goldrick, M., & Blumstein, S. E. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes*, *21*, 649–683.
- Hennessey, N. W., & Kirsner, K. (1999). The role of sub-lexical orthography in naming: A performance and acoustic analysis. *Acta Psychologica*, *103*, 125–148.
- Kawamoto, A. H., Liu, Q., Mura, K., & Sanchez, A. (2008). Articulatory preparation in the delayed naming task. *Journal of Memory and Language*, *58*, 347–365.
- Kello, C. T., Plaut, D. C., & MacWhinney, B. (2000). The task-dependence of staged versus cascaded processing: An empirical and computational study of Stroop interference on speech production. *Journal of Experimental Psychology: General*, *129*, 340–360.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–75.
- Li, X., Zhou, P., & Aruin, A. S. (2007). Teager–Kaiser energy operation of surface EMG improves muscle activity onset detection. *Annals of Biomedical Engineering*, *35*, 1532–1538. doi:10.1007/s10439-007-9320-z
- McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*, 287–330. doi:10.1037/0033-295X.86.4.287
- McMillan, C. T., & Corley, M. (2010). Cascading influences on the production of speech: Evidence from articulation. *Cognition*, *117*, 243–260.
- Possamaï, C. A., Burle, B., Osman, A., & Hasbroucq, T. (2002). Partial advance information, number of alternatives, and motor processes: An electromyographic study. *Acta Psychologica*, *111*, 125–139.
- Rastle, K., Croot, K. P., Harrington, J. M., & Coltheart, M. (2005). Characterizing the motor execution stage of speech production: Consonantal effects on delayed naming latency and onset duration. *Journal of Experimental Psychology: Human Perception and Performance*, *31*, 1083–1095.
- Rastle, K., Harrington, J., Coltheart, L., & Palethorpe, S. (2000). Reading aloud begins when the computation of phonology is complete. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1178–91.
- Roelofs, A. (2004). Seriality of phonological encoding in naming objects and reading their names. *Memory & Cognition*, *32*, 212–222.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662. doi:10.1037/0096-3445.121.1.15
- Theios, J., & Amrhein, P. C. (1989). Theoretical analysis of the cognitive processing of lexical and pictorial stimuli: Reading, naming, and visual and conceptual comparisons. *Psychological Review*, *96*, 5–24.