



**HAL**  
open science

## Interaction between articulatory gestures and inner speech in a counting task

Ali Hadian Cefidekhanie, Christophe Savariaux, Marc Sato, Jean-Luc Schwartz

► **To cite this version:**

Ali Hadian Cefidekhanie, Christophe Savariaux, Marc Sato, Jean-Luc Schwartz. Interaction between articulatory gestures and inner speech in a counting task. *Journal of the Acoustical Society of America*, 2014, 136 (4), pp.1869-1879. 10.1121/1.4893910 . hal-01073654

**HAL Id: hal-01073654**

**<https://hal.science/hal-01073654>**

Submitted on 10 Oct 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# **Interaction between articulatory gestures and inner speech in a counting task**

Ali Hadian Cefidekhanie, Christophe Savariaux, Marc Sato, Jean-Luc Schwartz<sup>(1)</sup>

GIPSA-Lab, Speech and Cognition Department,  
UMR 5216, CNRS – Grenoble University – France

**Suggested running title:** Articulatory gestures and inner speech

## Abstract

Interaction between covert and overt orofacial gestures has been poorly studied apart from old and rather qualitative experiments. The question deserves special interest in the context of the debate between auditory and motor theories of speech perception, where dual tasks may be of great interest. It is shown here that dynamic mandible and lips movement produced by a participant result in strong and stable perturbations to an inner speech counting task that has to be realized at the same time, while static orofacial configurations and static or dynamic manual actions produce no perturbation. This enables the authors to discuss how such kinds of orofacial perturbations could be introduced in dual task paradigms to assess the role of motor processes in speech perception.

Suggested PACS Classification numbers

Main section: 43.70

Detailed classification: 43.70.Bk, 43.70.Jt, 43.71.An

Keywords: perceptuo-motor link, double task, inner speech, orofacial gestures, mental counting, perturbation

# I. Introduction

## A. Searching for a motor modulation in speech perception tasks

### *1. The relationship between perception and action in speech communication*

Thanks to cognitive neuroscience discoveries in the last 20 years, the debate between auditory (e.g. Diehl et al., 2004), motor (e.g. Liberman & Whalen, 2000) and perceptuo-motor (Schwartz et al., 2012) theories of speech perception has entered a new stage. A large body of neuroimaging data highlights a cortical network connecting brain regions in the posterior frontal lobe, the parietal operculum and the posterior temporal lobe, providing a natural support to the association between auditory and motor speech representations (Hickok & Poeppel, 2004, 2007). A key question concerns the involvement of this so-called “dorsal” network in online communication processes.

While motor and premotor areas appear activated in various perceptual processing tasks (e.g. Fadiga et al., 2002; Watkins et al., 2003; Wilson et al., 2004; Alho et al., 2012; Grabski et al., 2013a), these activations do not, ipso facto, demonstrate the existence of a functional role of the motor system in perceptual processing. To assess whether there is indeed a functional role of motor processes in speech perception, one solution consists of trying to find a way to modulate the accessibility to production capacities and to examine whether this modulation affects comprehension.

### *2. Do motor perturbations modify speech perception?*

The first type of ‘modulation’ is provided by neurological data about aphasic patients. In their review of speech communication deficits, Hickok and Poeppel (2004, 2007) claim that there is a double dissociation between what they call ‘perception’ referring to

sublexical tasks (e.g. syllable discrimination) which would involve motor areas and the dorsal network, and 'comprehension' tasks involving lexical access and displaying no significant degradation in the case of frontal lesion of the premotor or motor areas. However, these sets of data remain complex and difficult to interpret, and provide only partial and limited tests of the hypothesis of a functional role of motor centers in comprehension mechanisms.

This is why researchers attempt to apply temporary (localized and of course reversible) micro-perturbations on specific regions of the frontal cortex (mainly on the motor cortex, the premotor cortex and Broca's area). Repetitive transcranial magnetic stimulation (rTMS) on the ventral premotor cortex produced small but significant modifications on phoneme identification in noise (Meister et al., 2007) or involving prior phonemic segmentation (Sato et al., 2009). TMS application at the level of specific articulators (lip vs. tongue) in the primary motor cortex produced selective changes in phonetic categorization or discrimination of phonemes in noise in relation to the involved motor area (d'Ausilio et al., 2009; Möttönen & Watkins, 2009), and the same kind of effect was obtained by used-induced motor plasticity (Sato et al., 2011). Disruption of the lip representation by TMS even produced an effect on the electro-encephalographic responses to auditory changes in the absence of acoustic noise (Möttönen et al., 2013). Finally, a recent study on selective adaption to speech also showed that TMS applied over regions of the dorsal pathway (the left ventral premotor cortex, supramarginal gyrus and posterior superior temporal gyrus) enhanced the auditory adaptation effect (Grasbki et al., 2013b). Thus, such temporary perturbations may show weak but consistent effects on some speech perception tasks, but they are indirect and difficult to monitor and assess precisely.

## **B. The dual-task paradigm**

### ***1. Principle***

Another way to assess the role of action in perception involves the dual-task paradigm, in which subjects perform a motor task concurrently with a perception task, to assess whether specific effects of the motor task could reveal a possible role for action in perception. The effect of manual gestures on visual tasks has been displayed in various situations, such as identification of the direction of a visual stimulus (in)consistent with the direction of the associated hand movement (Müsseler & Hommel, 1997), perception of the direction of motion in ambiguous stimuli (Wohlschlaeger, 2000), or perceptual judgment on the mass of a given object while lifting a more or less heavy similar object (Hamilton et al., 2004).

The dual-task paradigm has been extensively used by Baddeley and colleagues in the study of verbal working memory (Baddeley, 2003). Subjects performing various kinds of orofacial gestures display impaired ability to retain a list of speech items in short-term memory (Murray, 1968). This ‘articulatory suppression’ process is interpreted within the phonological loop model in which a mental rehearsal process allows for retention of the language stimuli in working memory, beyond the duration of phonological storage estimated around 2s (Baddeley & Hitch, 1974). The interpretation is that the perturbing task (for example continuously repeating a word) prevents or impairs mental rehearsal and therefore significantly reduces storage capacity. Dual tasks also produce a significant decrease in the verbal transformation effect (Reisberg et al., 1989) and perturb syntactic processing of complex sequences (Rogalski et al., 2008).

### ***2. Perceptual-motor content of articulatory suppression***

The dual-task paradigm has seldom been used in the study of phonetic decoding mechanisms in spite of its potential importance for the debate between motor and auditory theories of speech perception. The reason is likely the difficulty in accurately characterizing and controlling motor perturbations. Most experiments on articulatory suppression involved the repetition of speech stimuli aloud, whether meaningful or not (typically mono- or bi-syllables). This induces an articulatory and an auditory component, both subject to possible interference with the target perceptual task. A series of ingenious experiments by Gupta & MacWhinney (1995) confirmed that both components were involved in articulatory suppression in verbal working memory. Evidence for auditory interference was provided by the fact that articulatory suppression was decreased for silent vs. aloud speech production. Pure articulatory interference was also displayed since articulatory suppression was stronger than a mere perturbation by concomitant sound stimuli (irrelevant speech effect).

Therefore silent or internal production (inner speech) seems more appropriate to avoid auditory interference. However, silent speech production also generates auditory imagery (what Gupta & MacWhinney refer to as 'speech inside the head'), through efference copy. If a dual-task experiment involving speech perception and silent speech production changes the results of the perceptual task, we cannot easily distinguish between (1) the perceptual system's recourse to motor skills for categorization (Figure 1a), which would be modulated by the engagement of the motor system in the perturbing task (Figure 1b), and (2) the auditory imagery interference generated by the motor perturbation on the target stimulus to process perceptually (Figure 1c). Interference with auditory imagery is suggested by Sams et al. (2005) in an experiment involving the silent production of a syllable "pa" or "ka" concomitant with auditory categorization of a noisy syllable "pa" and "ta", with effects similar to those of audiovisual fusion (increase of the percentage of

correct answers in the case of concordant perceived and produced stimuli, decrease in the case of discordant stimuli, with the generation of an analog to the McGurk effect, McGurk & MacDonald, 1976) (see also Kauramäki et al., 2010; Sato et al., 2013). The interpretation by Sams and coll. that the effect is due to auditory imagery (Fig. 1c) is based on neurophysiological data showing that silent articulation modifies the activity of the auditory cortex (e.g. Houde et al., 2002), though the concurrent interpretation corresponding to Fig. 1b cannot be dismissed.

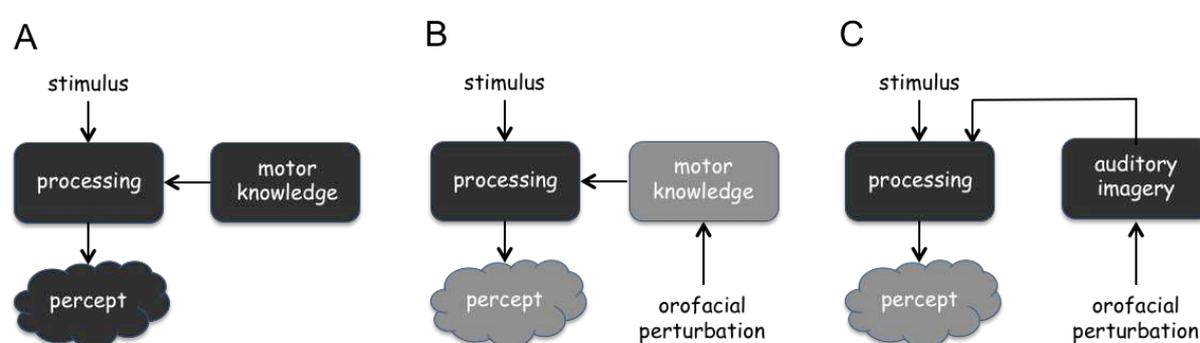


Figure 1

This is why it seems preferable to use a task mobilizing the orofacial system without speech production (even covert) per se. This was the rationale in a study by Reisberg et al. (1989) on the verbal transformation effect, showing a decrease in the number of reported verbal transformations by perturbing tasks such as locking the mandible, keeping the lips pressed together and the tongue stuck to the palate, or chewing gum. Convergent evidence from neuroimaging confirms that while auditory areas in the cortex are involved in speech production (e.g. Bohland & Guenther, 2006), they are not activated (or in any case significantly less so) by the production of elementary orofacial

gestures such as opening the mandible, protruding the lips or withdrawing the tongue (Grabski et al., 2012).

Therefore, the production of such orofacial gestures should engage the production system and thus potentially modulate its efficiency in the perceptual task (Figure 1b) without producing auditory imagery by means of efference copy (Figure 1c).

### ***3. Finding efficient articulatory gestures for articulatory suppression***

It remains unknown if some articulatory gestures are more efficient than others in engaging the speech production system, and thus play a possible role in modulating perception. This is the focus of the present research. To address this question we used a dual-task paradigm with two motor tasks, exploring how various kinds of orofacial gestures might perturb the generation of inner speech. The rationale was that inner speech draws on a cortical network partly common to that of overt speech production (Yetkin et al., 1995), and therefore involves the network of what we call ‘motor knowledge’ in Figure 1. If an overt orofacial task can effectively mobilize the speech production system, then it should significantly perturb inner speech (Figure 2).

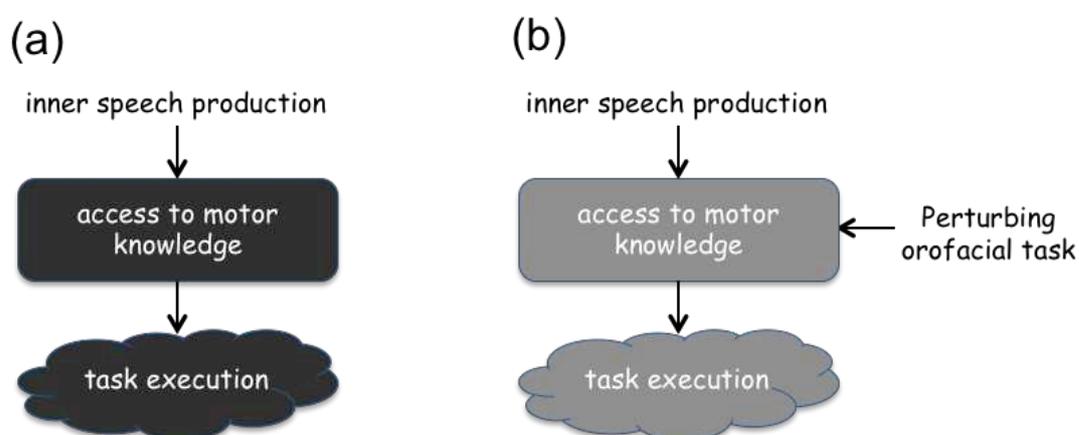


Figure 2

The outcome is not certain and has been hotly debated in the past. In his study of inner speech during silent reading, Pintner (1913) puts forward the opposing positions of Stricker (1880, cited in Pintner, 1913) and Paulhan (1886, cited in Pintner, 1913), the former considering that one cannot have “the idea of the ‘b’ sound without the feeling of a muscular movement or an innervation of the lips” and the latter claiming that he was able to formulate the idea of one vowel while uttering another. Jeannerod (2003) recalls that “Authors of the time (e.g. Binet, 1886) claimed that mental images in general resulted from excitation of the same cerebral centers as the corresponding actual sensation (...) for example, it was shown to be impossible for a subject to generate the image of pronouncing the letter /b/ if he kept the mouth wide open: this was because, supposedly, the motor system cannot be engaged in two contradictory actions at the same time.”

Dual-task paradigms involving an overt and a covert task have been used with manual gestures (for example in tasks of mental rotation, Wexler et al., 1998; Wohlschläger & Wohlschläger, 1998) or locomotion (Kunz et al., 2009). As far as we know, they have never been directly used and quantitatively tested in speech. In this paper, we explored different types of perturbation (mandible or lips movements in static vs. dynamic, fast vs. slow modes, compared with hand movements used as a control) in an inner speech task consisting of mental counting. Counting time is used as the index of potential perturbation, with the hypothesis that the stronger a perturbation, the longer the counting time.

In the next section we present the material and methods used for this study. Results are described in Section III and discussed in Section IV, before a conclusion is presented in Section V.

## **II. Material and methods**

### **A. Paradigm and task**

The study involved doing two things at once, in order to explore the interaction between a target task – mentally counting – and a perturbation task – producing overt articulatory or manual gestures.

The target task consisted of a series of two consecutive countings from 1 to 30 as fast as possible, in 11 different conditions (see Table 1). Counting was done aloud in one condition and mentally in the other ten. Out of these ten conditions, one was unperturbed, while the other nine involved a motor task perturbation. The perturbing motor task consisted of producing, at the same time as mental counting, a specific action with one of three effectors: left hand, mandible or lips. Three types of action were required: static (keeping the hand open, keeping the mandible low, keeping the lips protruded) or dynamic-cyclic (opening and closing the hand, lowering and raising the mandible, protruding and stretching the lips) at two possible speeds: slow or fast (0.5 cycle vs. 1 cycle per second). For each participant, the 11 conditions were presented in a random order in five full consecutive blocks.

Table 1 – The 11 conditions of the experimental paradigm

1	No hand/face movement	Overt counting (loudly)	Mental counting
2		Covert counting	
3	Static (no movement)	Hand	
4		Jaw	
5		Lips	
6	Dynamic (slow rate)	Hand	
7		Jaw	
8		Lips	
9	Dynamic (fast rate)	Hand	
10		Jaw	
11		Lips	

For each test, the procedure included the following sequence of steps for the subject (Figure 3):

- (1) read on the screen the appropriate task among the 11 possible conditions;
- (2) click a button with their right hand to start the process;
- (3) if required, launch the perturbation task; if static, produce the appropriate configuration of the left hand (open), mandible (low) or lips (protruded); if dynamic, synchronize the movement of either their left hand or their mandible or their lips to the pace of a red cross displayed on the screen, flashing at a rate of 0.5 cycles per second (if slow) or 1 cycle per second (if fast);
- (4) once the perturbing task (static or dynamic) is launched, keep it active and close the eyes to concentrate so as not to be disturbed by the cross;

(5) press on a key with their right hand at the exact time of starting the first round of counting from 1 to 30;

(6) press again with the right hand at the end of the first cycle to start the second cycle;

(7) press for the last time with their right hand to signal the end of counting and proceed to the next test.

A short training phase (for the 11 conditions presented consecutively) prepared the subjects to understand and perform the task.

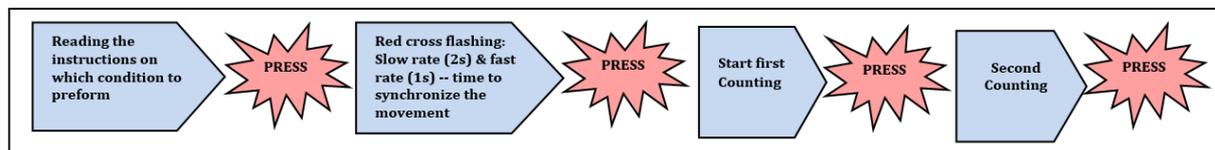


Figure 3

## B. Setup and subjects

10 native right-handed French subjects (5 female and 5 male), aged between 22 and 36 years old (mean 26 and a half), without visual or auditory problems, participated in the experiment (after giving their informed consent).

The experiment was conducted in an anechoic chamber, with the Presentation® software ([www.neurobs.com](http://www.neurobs.com)). To check that the perturbing task was correctly performed, and to assess possible perturbations of the perturbing task by the target task in return, the subjects' lips were made up with blue lipstick, and one small round blue sticker was stuck on their chin, another one on the inside of their left hand and a third one on the last joint of their middle finger (Figure 4). A Chroma-Key system, which transforms blue areas into pure black, automatically detects the corresponding zones, which allows to

quantitatively analyze the movements of the lips, mandible (Lallouache, 1990) or hand/finger (Heracleous et al., 2010).



Figure 4

## C. Analyses

### 1. Target task

The inner counting task constituted the target task in this study, and it was evaluated through the measurement of counting duration. For each of the 10 subjects, each of the 11 conditions and each of the 5 repetitions, there were two consecutive countings from 1 to 30, and the duration of each counting was calculated from the difference between the time of successive button pressings (first counting: time delay between steps (5) and (6) in Section II.A; second counting: time between steps (6) and (7)).

The hypothesis was that counting times in the 9 conditions of inner counting with perturbation would be higher than in the inner counting with no perturbation condition. The aloud counting condition (“overt” from now on) was used as a reference to check that the mental counting task without perturbation was well performed, on the assumption that counting time should be pretty much equal between real and imagined actions (Jeannerod, 1995). Hand perturbation was used as a control on the assumption that it would only slightly slow down mental counting (because of the simple effect of a dual-task paradigm) while orofacial perturbations (mandible or lips) should produce larger perturbations because of interference in the same action system (the orofacial system, engaged in both mental counting and lips or mandible perturbation). We had no strong *a priori* hypothesis concerning the most perturbing effect between static, dynamic fast, or dynamic slow perturbations.

The repetition of the counting action twice consecutively enabled us to assess whether the perturbation would be stronger at the beginning and decrease later due to progressive adaptation to the task: in this case, there would be less perturbation during the second counting than the first one. In the same vein, it could be predicted that perturbation would decrease along the 5 blocks.

## ***2. Perturbation task***

Thanks to the blue make-up on the lips and the blue stickers on the chin, hand and finger, we were able to follow lips, jaw and hand movements over time. In Figure 5, we have displayed a typical trajectory for one articulator (lips). Each trajectory was analyzed in terms of a succession of cycles, of which the period was automatically extracted as displayed in the Figure. For lips and jaw, additional measurements could be done on the movement amplitude for each cycle. This was not feasible for hand/finger, for which the

information disappears in some parts of the trajectory because of the occlusion of the stickers.

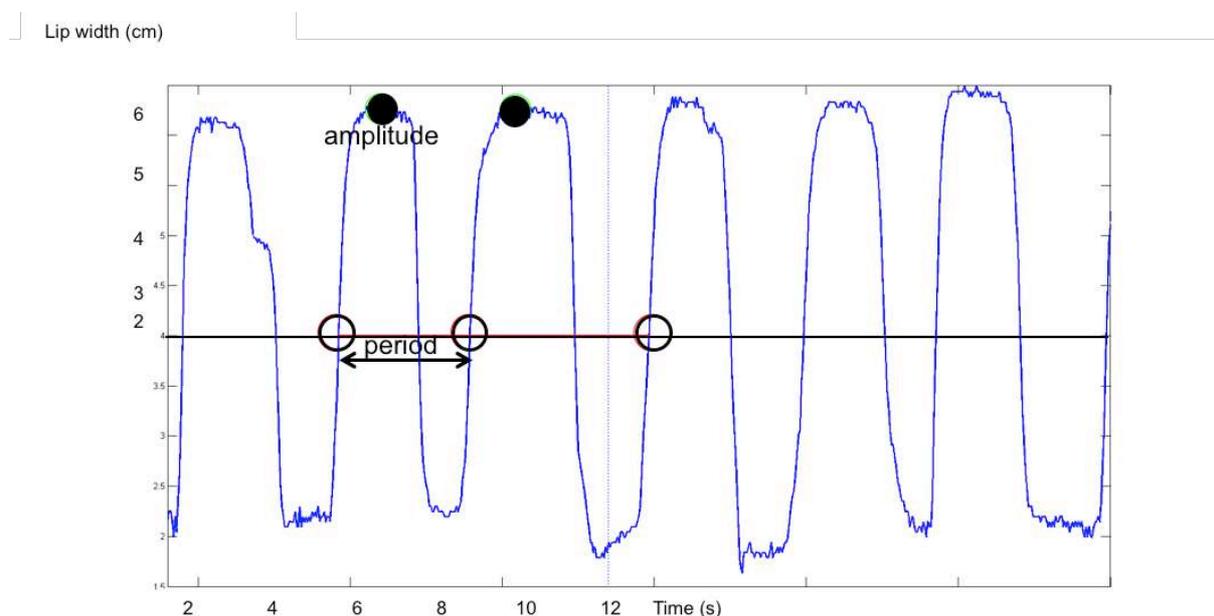


Figure 5

For each trial (that is for each subject and each condition), the time when the subject closed the eyes gave the information about the time when internal counting began, which enabled us to define two portions in the trajectory, “before” (corresponding to the perturbing task alone) and “after” (corresponding to the perturbing task performed at the same time as the inner counting task). This enabled us to see if the perturbing task was correctly realized and also to assess whether the target task (inner counting) perturbed the perturbing task in return. Notice that the instant of the first button pressing for inner counting would have provided a more precise estimation of inner counting onset, but this information was not available on the video track. However, it was checked during the experiment that there was a good overall correspondence between eye closure and first

button pressing, the more so considering that analyses of means and variances of cycle periods “before” and “after” do not require a precision better than one or two seconds for positioning the boundary.

### **III. Results**

#### **A. Inner speech**

##### ***1. Global results***

For each subject, we extracted the counting times (T) for each of the 11 conditions (C), each of both counting essays (E) and each of the five blocks (B). An analysis of the variance (ANOVA) with repeated measures on three factors (C, E, B) was performed with Greenhouse-Geisser correction in cases of violation of the sphericity assumption. This analysis revealed an effect of condition C [ $F(10, 90) = 20.56, p < 0.001$ ] and of essay E [ $F(1, 9) = 14.93, p < 0.005$ ], with no other effect, neither of the block nor of any interactions of two or three factors.

The effect of condition is summarized in Figure 6. The overt counting condition produces the shortest counting time (6.25 s), while the inner counting without perturbation condition (‘covert’) takes approximately 1.5 s longer (7.77 s), contrary to the claim by Oppenheim & Dell (2010) that “inner speech is faster than overt speech”. However the difference is not significant according to a post-hoc test with Scheffé correction. Dynamic orofacial gestures of the mandible or lips, whether fast or slow, produce the longest counting times. We will return to the statistical analysis of these differences later.

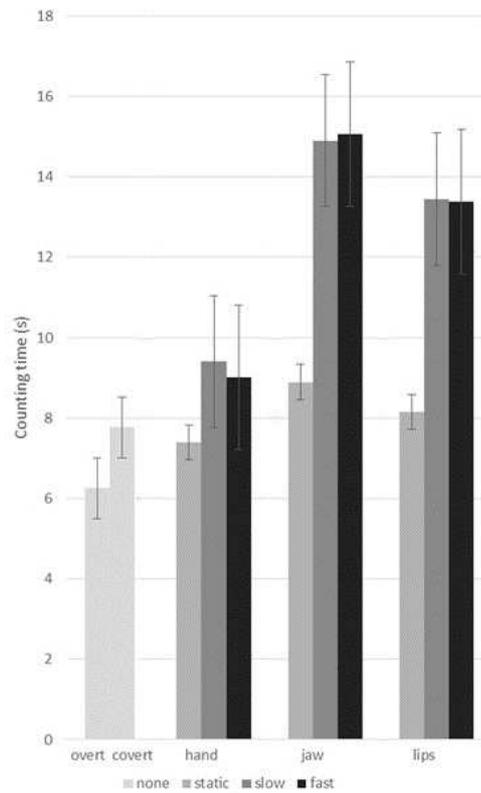


Figure 6

## ***2. Stability of the perturbation effect with essay and block***

Regarding the essay factor (E), the average counting time for the first essay was significantly shorter than the one for the second test (10 s against 10.7 s). An observation of this difference for the 11 conditions revealed a tendency for the difference to increase with the duration of the first count. The correlation between the time of the first essay and the difference between the two essays is significant [ $r^2 = 0.56$ ,  $t(9) = 3.39$ ,  $p < 0.01$ ], with an increase of 7% in counting time from the first to the second essay. This suggests a general trend for slowing down along a given counting task whatever the condition and the subject.

The important point is that this slowdown, which is proportional to the duration of the counting, does not reflect the ability of the subjects to cope more easily with the perturbation from one essay to another (this would result instead in a decrease in the duration of the second essay compared with that of the first one, especially in the case of perturbation). So, the effect of the perturbation does not decrease from the first essay to the second one. The lack of any significant effect of block alone or in interaction shows that the perturbation effects are stable throughout the whole experiment. In the following, counting times will therefore be averaged over counting essay (E) and block (B).

### ***3. Effect of the type of perturbation***

In order to analyze the effect of different perturbations more precisely, we conducted a second analysis of the effects of the three modes of perturbation (M, for static vs. slow dynamic vs. fast dynamic) and the three involved gestures (G, either hand, mandible or lips), by performing a repeated measures ANOVA with two within-subjects factors (M and G) on the 9 conditions in dual tasks. The dependent factor for each subject and condition was the counting time T averaged over the 2 essays and the 5 blocks. Both factors are significant [factor M:  $F(2, 18) = 17.476$ ,  $p = 0.002$ ]; factor G:  $F(2, 18) = 22.36$ ,  $p < 0.001$ ] as well as their interaction [ $F(2, 18) = 11.24$ ,  $p = 0.001$ ]. The post hoc tests (with probability  $< 0.05$  with Scheffé correction) show that:

- The static perturbation produces significantly shorter counting times than the dynamic perturbation, either slow or fast (with no significant difference between the two speeds);
- The hand movement produces significantly shorter counting times than the two orofacial movements, mandible and lips, which do not differ;

- If we take the static hand perturbation as a reference, the static perturbations in the 3 types of gestures as well as the 2 dynamic perturbations with hand movements do not produce significantly differing counting times, whereas the slow and fast dynamic perturbations, for both orofacial articulators (mandible or lips) produce significantly longer counting times, not significantly different from each other.

In summary:

**(static hand/mandible/lips = dyn. slow/fast hand) < (dyn. slow/fast mandible/lips)**

## **B. Overt gestures**

For each condition where there was a dynamic perturbation task (slow or fast cycles of the jaw, lips or hands), for each subject and each trial, we computed the average cycle period respectively before and after the time when the subject closed their eyes. Then we computed the mean and relative standard deviation (that is standard deviation divided by the mean) of these average periods over all subjects and trials to assess whether the perturbing task was correctly performed. In Figure 7 we have displayed these mean and relative standard deviations for the three gestures and the two speeds. Two major observations can be drawn from this figure.

First, the task is realized fairly accurately, as shown by the fact that the mean cycle periods are around 2s for the slow condition and around 1s for the fast condition for the three types of gestures.

Secondly, the difference between “before” and “after” eye closure is not so large in terms of mean cycle modifications (the mean cycle being actually quite stable in the fast

conditions, and slightly less in the slow conditions with a period decrease between 5 and 10%) than in terms of increasing dispersion, as displayed by the relative standard deviation values which are about twice larger “after” than “before” in the fast conditions, and three times larger in the slow conditions. Of course, this is not surprising considering that when subjects close their eyes they can no more monitor the cycling pace and hence they become more variable, though they keep the requested mean rhythm as displayed by the stability of the mean period. We shall evaluate in the next section if modifications differ from one type of gesture to another.

Finally, analysis of amplitudes for lips and jaw display a small reduction in amplitude (around 3 to 5%) from “before” to “after” eye closure, in all conditions (for the two articulators and the two speeds).

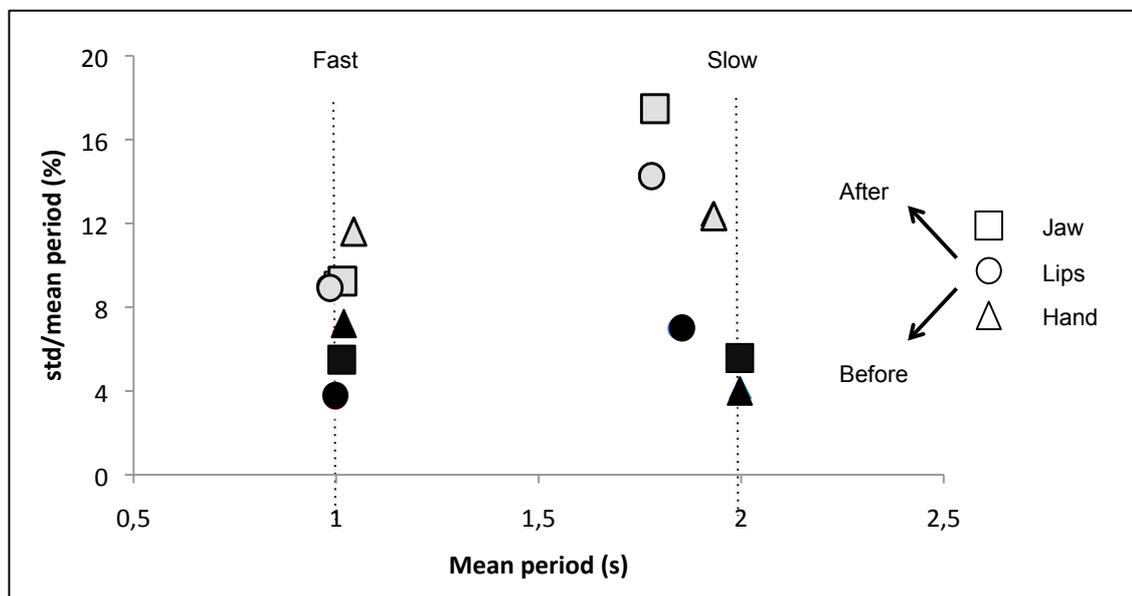


Figure 7

### **C. Comparison between the target counting task and the perturbing gesturing task**

Finally, we attempted to compare the way the target task (i.e. inner counting) was perturbed by the dynamic perturbing task (i.e. slowly or quickly gesturing in cycles with the jaw, lips or hand) with the way the perturbing task was perturbed in return by the target task, and how the type of gesture interfered in this process. Information about the target task alone was provided by condition 2 (“covert counting”, in Table 1). Information about the gesturing task alone was provided by the data gathered before eye closing. Information about interfering tasks (counting with the gesturing perturbation) was provided by counting time on one hand, cycling period on the other hand.

Therefore, for each subject and each gesturing condition (two speeds and three types of gestures), we computed both:

- (1) the size of the difference in counting time from the target task alone to the double task situation (variation in Target task due to Perturbation task)

$$\text{Diff}(\text{Target}/\text{Perturbation}) = \text{abs}(\text{counting}(\text{Target alone}) - \text{counting}(\text{Target} + \text{Perturb.}))$$

- (2) and the size of the difference in cycling period from the perturbation task alone to the double task situation (variation in Perturbation task due to Target task)

$$\text{Diff}(\text{Perturbation}/\text{Target}) = \text{abs}(\text{period}(\text{before eye closure}) - \text{period}(\text{after eye closure}))$$

Then we assessed how absolute differences, supposed to provide an index on how one task modified the realization of the other task, depended upon the first task, the type of gesture and the cycling speed. This was done through a repeated-measures ANOVA with

three within-subject factors: Task (Target/Perturbation vs. Perturbation/Target), Gesture (jaw, lips, hand) and Speed (slow, fast).

There were two basic assumptions in this comparison. Firstly, the target task should be more perturbed than the perturbation task: the Diff(Target/Perturbation) values should be larger than the Diff(Perturbation/Target) values. Secondly, jaw and lips should result in larger interference values than hand. There were no strong assumptions about speed.

All results are displayed in Figure 8. The ANOVA was done on logarithms of the Diff values to better ensure Gaussian distributions (Diff values being always positive by definition), with Greenhouse-Geisser correction in cases of violation of the sphericity assumption. It shows that:

- There is an effect of the Task ( $F(1,9) = 49.67, p < 0.001$ ) with much larger effects of “Perturbation” on “Target” than the inverse, which was quite expected. The Diff values displayed in Figure 8 are typically 3 to 6 times larger when the hand is involved, and 10 to 20 times larger when the jaw or lips are involved. Notice that while Diff(Target/Perturbation) does represent the difference between a single and a double task, Diff(Perturbation/Target) cumulates the effect of Target on Perturbation with the effect of eye closure and progressive drift from the visual pace, which is likely to be the major factor responsible for the modification of the gesture cycling task. Hence the real difference is actually largely underestimated by the present analysis.
- There is both an effect of Gesture ( $F(2,18) = 27.88, p < 0.001$ ) and of Gesture and Task interaction ( $F(2,18) = 4.79, p = 0.02$ ). Post-hoc analyses show that the difference in Gesture is between hand and the (jaw, lips) group, and that the difference only comes from the Target/Perturbation condition (which corresponds

to the result in Section II.A.3), while there is no significant difference between gestures in the Perturbation/Target condition. This is very important. Indeed, one might think that the lower perturbation of the counting task due to the hand could arise from the fact that the hand gesture was less well realized and hence more perturbed than the jaw or lips gesture. In actual fact, the trend, according to Figure 8, is rather the reverse: it seems that the perturbation of jaws and lips is larger than the perturbation of hand, which confirms that there is more interference between counting and the orofacial system, possibly in both directions (from one task to the other and vice versa) than between counting and the manual system.

- There is also an effect of Speed ( $F(1,9) = 28.72$ ,  $p < 0.001$ ), and the interaction between Speed and Task is close to significance ( $F(1,9) = 4.78$ ,  $p = 0.057$ ). It appears that the difference between conditions “fast” and “slow” appear mainly in the Perturbation/Target condition, where the effect is larger in the slow than in the fast condition. On the other hand, there does not seem to be any difference between the two speed conditions in the Target/Perturbation condition, which is coherent with the lack of a speed effect in the analysis in Section III.A.3.

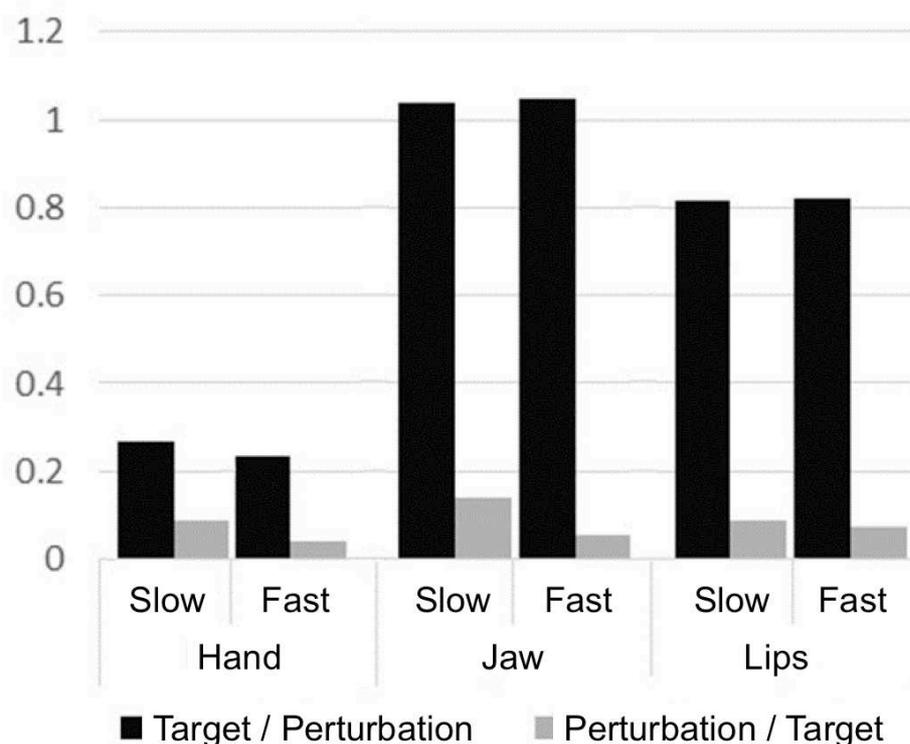


Figure 8

## IV. Discussion

### A. Lack of effect of static perturbations on inner speech

The first important result is that a static perturbation seems ineffective, regardless of the articulator: counting times were not significantly different regardless of the type of the effector involved, whether manual or orofacial, and they were similar to counting times in cases without perturbation (see Figure 6). This contradicts earlier claims by Stricker or Binet: inner speech is not significantly perturbed by static orofacial positioning, whatever the type. A study by Tuomainen et al. (2002) cited by Sams et al. (2005), seems to confirm this point. In that study, Finnish participants were tested on the categorization of acoustic vowels on the /Q/-/ø/ continuum in Finnish, while silently articulating /Q/

or /ø/. Effects of silent articulation were displayed, e.g. silent articulation of /Q/ shifted the phoneme boundary significantly toward /ø/. However, no shift was displayed when subjects were instructed to position their articulation system as if they would say /Q/ or /ø/, without proper articulation.

A possible interpretation can be found in the literature on speech production perturbations by e.g. bite-block (Lindblom et al., 1977) or lip-tube (Savariaux et al., 1995). These studies show that subjects were largely able to adapt to a static perturbation. Moreover, if adaptation is not complete (which is the case in the lip-tube paradigm), possible induced modifications in auditory imagery could be partially dealt with in a motor-to-auditory efferent copy process in the context of predictive coding models of speech production (Hickok, 2012). On the contrary, a dynamic perturbation such as the one induced by a cyclic movement of the lips or jaw results in constantly varying sensory-motor configurations, which largely complicate the possibility of dealing with the perturbation over time.

## **B. Mandible and lips dynamic gestures provide efficient and stable perturbations**

The dynamic perturbations produce the desired effect: while a hand perturbation minimally changes the counting time compared to the condition without perturbation, the two orofacial perturbations produce a significant perturbation. If we take the dynamic hand condition as a reference for a dual-task condition, the counting time goes from 7.5s to about 14s for the 4 dynamic orofacial conditions, which represents an increase of almost 100%. The fact that the two paces of dynamic orofacial perturbation produce

similar effects may seem surprising. We could have expected that the mental production would synchronize with the perturbation and therefore would slow down with the drop in speed, or on the contrary that a slow speed would leave more room for a parallel mechanism of production and therefore produce less perturbation. We now need to find out what happens with other speeds in order to determine if all orofacial perturbations give rise to similar results or whether the effects of locking to certain speeds could affect the performance in the dual-task paradigm.

The fact that an orofacial perturbation provides more effect than a manual perturbation on an inner task (counting), which is intrinsically orofacial, cannot appear surprising, considering the large literature on dual-tasks paradigms (see Section I.B.3). However, a possible argument could be raised, according to which orofacial gestures would be intrinsically more complex than manual gestures. They would hence be more demanding in terms of planning/representation/execution than manual gestures and in consequence would produce more perturbation whatever the target task<sup>(2)</sup>. To test this assumption, we performed a control experiment involving a target task based on inner manual gestures instead of inner orofacial gestures. When a dual-task paradigm is applied to this target task, it appears that orofacial gestures are no longer more perturbing than manual gestures (see Appendix). This discards the possibility that orofacial gestures would be intrinsically more perturbing than manual gestures whatever the inner target task. It confirms that the perturbation of counting time associated with orofacial gestures is specifically due to the fact that the two actions (one overt, the perturbation task, and one covert, the target task) are applied on the same action system (the orofacial system) and therefore compete for the same cognitive/neural resources.

### **C. A guideline for future dual tasks experiments**

The major result of this study is that we can achieve a perturbation of inner speech by orofacial movements (opening/closing cycles of the mandible, or protruding/stretching cycles of the lips), which should not generate auditory images at the same time by efference copy. The fact that these perturbations are quite stable from one counting essay to another shows that the perturbation continues unabated for at least 25 to 30 seconds. The lack of differences along the five blocks confirms the stability of the perturbation effectiveness. The fact that similar effects are obtained for the lips and the mandible and for the 2 speeds provides an additional favorable factor for the design of a dual-task paradigm in the study of motor effects in speech perception. Indeed, taking the dynamic hand condition as a control for the dual-task, it may be possible to vary the orofacial effector (mandible or lips) and the speed throughout an experiment to avoid adaptive mechanisms that could reduce or remove the perturbation effect. The appeal of the perturbation task is that it is easy to monitor and control for the experimenter, and easy to produce by the participants, as displayed by the rather stable pattern of cyclic characteristics in Figure 7.

Therefore, there is now room, thanks to the present data, for a well-controlled dual task experiment assessing the role of motor knowledge in speech perception.

## **V. Conclusion**

In a carefully controlled experimental setup, a dual task combining inner counting (Target) and hand or orofacial actions (Perturbation) has been explored and analyzed. It appears that static configurations of the jaw, lips or hand do not modify inner counting

time, hence provide no apparent perturbation. This is also the case with dynamic hand movements regardless of their speed. On the other hand, slow or fast cyclic movements of the lips or jaw significantly increase inner counting time, by almost 100%. The perturbation is stable over time and the perturbing task is easy to monitor and control. Since mono-articulator gestures are unlikely to produce auditory imagery effects, it is proposed that such articulatory dynamic movements should provide a good basis for dual task experiments in speech perception, attempting to assess the role of the orofacial motor system in various kinds of experimental perception conditions.

## **Appendix – A control experiment assessing hand vs. jaw perturbations on an inner manual task**

### **A. Rationale**

To discard the possibility that orofacial gestures would be cognitively more demanding than manual ones and would hence produce more perturbations on any inner task, we designed a control experiment in which the target task was based on a sequence of hand gestures. The experiment was based once again on a dual-task paradigm with the same perturbation tasks with the hand and jaw as in the counting experiment.

### **B. Materials and Methods**

#### ***1. Paradigm and task***

The target task required a sequence of gestures to be realized with the right hand. It consisted of successively touching with the right hand the forehead, the chin, the left ear,

the right ear and the nose, and to repeat the sequence a second time in the same order. We designed this task to try to induce a behavior that, when realized in the covert condition, would be difficult to mentalize without motor imagery involving the hand. The sequence was repeated two times to obtain a sufficiently long achievement time (a few seconds) so that perturbation effects could be possibly displayed and measurable. As in the counting task, the subject pressed the key (with the left hand), began a first essay of two manual sequences, pressed again the key, realized a second essay of two manual sequences, and pressed the key a last time so that the duration of the two essays could be measured.

The target task was realized in 8 different conditions: overtly in the first one and covertly in the other seven. Among the seven covert tasks, one was unperturbed and the six others were realized together with another motor task (perturbation). The perturbing motor task consisted of producing, at the same time as the covert manual task, a specific action with one of two effectors: either the right hand or the mandible. Three types of action were considered: static (keeping the hand open, or keeping the mandible low) or dynamic-cyclic (opening and closing the hand, or lowering and raising the mandible) at two possible speeds: slow or fast (0.5 cycle vs. 1 cycle per second). For each participant, the 8 conditions were presented in a random order in five full consecutive blocks.

The procedure was the same as the one described in the counting task (see §II.A and Figure 3) except that the participants were required to use their left hand to press the key all along the test, the right one being occupied by the perturbation. A short training phase (for the 8 conditions presented consecutively) prepared the subjects to understand and perform the task (particularly to learn to mentalize the sequence of manual gestures).

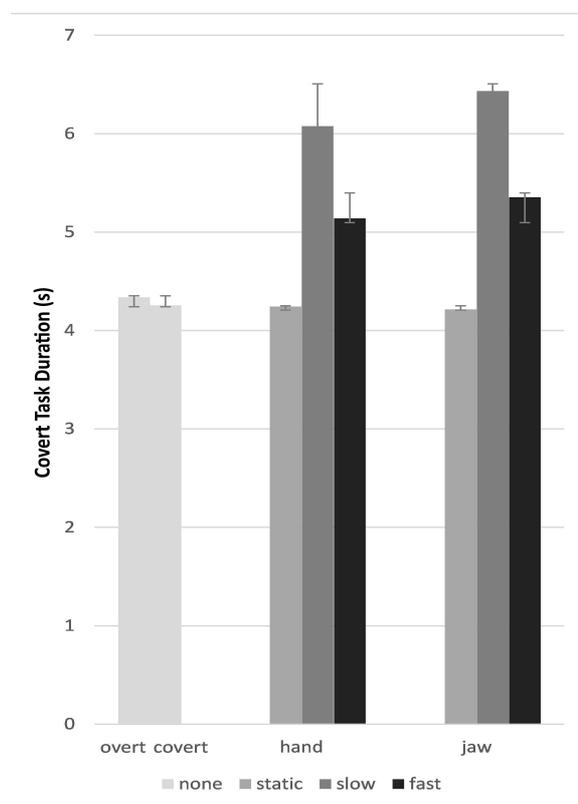
## ***2. Setup and subjects***

12 native right-handed French subjects (10 females and 2 males) participated in the experiment (after giving their informed consent). None of them had participated to the counting task.

The experiment was conducted in an anechoic chamber, with the Presentation® software (www.neurobs.com). To check that the perturbing task was correctly performed the subjects were video recorded but no analysis was done from these videos.

### C. Results

For each of the 12 subjects, each of the 8 conditions and each of the 5 repetitions, there were two consecutive essays (made of two manual sequences), and the duration of each essay was calculated from the difference between the times of successive button pressings. The effect of condition (averaged over subjects, repetitions and essays) is summarized in Figure A1.



## Figure A1

In order to analyze the effects of the three modes of perturbation (M, for static vs. slow dynamic vs. fast dynamic) and the two involved gestures (G, either hand or mandible), we performed a repeated-measures ANOVA with two within-subjects factors (M and G) on the 6 conditions in dual tasks. The dependent factor for each subject and condition was the covert task time T averaged over the 2 essays and the 5 blocks. While the speed factor was significant ( $F(2, 22) = 21,594, p < 0.001$ ), the gesture factor was not significant, nor was the interaction between gesture and mode. Post-hoc analysis on mode showed that the static condition was significantly quicker (and actually not different from the unperturbed covert condition, see Figure A1): once again a static dual task produces no perturbation. The two dynamic conditions produced significant perturbations, and the perturbation was significantly larger for the slow rhythm compared with the quick rhythm.

## Discussion and conclusion

Importantly, it appears that the perturbation due to dynamic gestures is the same for manual and orofacial gestures. Considering the fact that the target task was manual, we actually expected more perturbation with the hand than the mandible. The fact that we did not obtain such a difference is likely due to the involvement of the jaw in the required covert task, possibly because participants had to realize a sequence which led them produce some kind of unconscious counting. It is also possible, according to posterior debriefing of the subjects, that pointing towards the forehead, chin, ears and nose was actually done mentally by other actuators than the hand, e.g. the eyes or even the jaw, alone or in coordination with mental manual touch movements. Pointing with the eyes

would lead to no perturbation at all, either by the manual or by the oral gestures, which actually seemed to be the case for some subjects. Pointing with the jaw, in coordination with the hand, would lead to similar perturbations by manual and oral gestures.

However, this control task clearly discards the possibility that orofacial gestures would be intrinsically more perturbing than manual gestures whatever the inner target task.

## **Endnotes**

(1) Corresponding author ([jean-luc.schwartz@gipsa-lab.grenoble-inp.fr](mailto:jean-luc.schwartz@gipsa-lab.grenoble-inp.fr))

(2) We thank one of the reviewers for this interesting comment, which led us complete the control experiment presented in Annex 1.

## References

- ALHO, J., SATO, M., SAMS, M., SCHWARTZ, J.-L., TIITINEN, H., & JAASKELAINEN, I.P. (2012). Enhanced early-latency electromagnetic activity in the left premotor cortex is associated with successful phonetic categorization. *Neuroimage*, *60*, 1937-1946
- D'AUSILIO, A., PULVERMÜLLER, F., SALMAS, P., BUFALARI, I., BEGLIOMINI, C., & FADIGA, L. (2009). The motor somatotopy of speech perception. *Current Biology*, *19*, 1-5.
- BADDELEY, A.D. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience*, *4*, 829-39.
- BADDELEY, A.D., & HITCH, G. (1974). *Working memory*. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47-89). New York: Academic Press.
- BOHLAND, J.W., & GUENTHER, F.H. (2006). An fMRI investigation of syllable sequence production. *Neuroimage*, *32*, 821-841.
- DIEHL, R.L., LOTTO, A.J., & HOLT, L.L. (2004). Speech perception. *Annual Review of Psychology*, *55*, 149-179.
- FADIGA, L., CRAIGHERO, L., BUCCINO, G., & RIZZOLATTI, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, *15*, 399-402.
- GRABSKI, K., SCHWARTZ, J.-L., LAMALLE, L., VILAIN, C., VALLEE, N., BACIU, M., LE BAS, J.-F., & SATO, M. (2012). Functional MRI assessment of orofacial articulators: neural correlates of lips, jaw, larynx and tongue movements. *Human Brain Mapping*, *33*(10), 2306-2321.

- GRABSKI, K., SCHWARTZ, J.-L., LAMALLE, L., VILAIN, C., VALLÉE, N., BACIU, M., LE BAS, J.-F., & SATO, M. (2013a). Shared and distinct neural correlates of vowel perception and production. *Journal of Neurolinguistics*, *26*(3), 384-408.
- GRABSKI, K., TREMBLAY, P., GRACCO, V., GIRIN, L., & SATO, M. (2013b). A mediating role of the auditory dorsal pathway in selective adaptation to speech: a state-dependent transcranial magnetic stimulation study. *Brain Research*, *1515*, 55-65.
- GUPTA, P., & MACWHINNEY, B. (1995). Is the articulatory loop articulatory or auditory? Reexamining the effects of concurrent articulation on immediate serial recall. *Journal of Memory and Language*, *34*, 63-88.
- HAMILTON, A., WOLPERT, D., & FRITH, U. (2004). Your own action influences how you perceive another person's action. *Current Biology*, *14*, 493– 498.
- HERACLEOUS, P., BEAUTEMPS, D., & ABOUTABIT, N. (2010). Cued Speech Automatic Recognition in Normal-hearing and Deaf Subjects. *Speech Communication*, *52*, 504-512.
- HICKOK, G., & POEPEL, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition*, *92*, 67–99.
- HICKOK, G., & POEPEL, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*, 393–402.
- HICKOK, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*, 135-145.
- HOUDE, J.F., NAGARAJAN, S.S., SEKIHARA, K., & MERZENICH, M.M. (2002). Modulation of the auditory cortex during speech: an MEG study. *Journal of Cognitive Neuroscience*, *14*, 1125– 1138.
- JEANNEROD, M. (1995). Mental imagery in the motor context. *Neuropsychologia*, *33*, 1419–1432.

- JEANNEROD, M. (2003). The mechanism of self-recognition in humans. *Behavioral and Brain Research, 142*, 1–15.
- KAURAMÄKI, J., JÄÄSKELÄINEN, I.P., HARI, R., MÖTTÖNEN, R., RAUSCHECKER, J.P., & SAMS, M. (2010). Lipreading and covert speech production similarly modulate human auditory-cortex responses to pure tones. *Journal of Neuroscience, 30*, 1314–1321.
- KUNZ, B.R., CREEM-REGEHR, S.H., & THOMPSON, W.B. (2009). Evidence for motor simulation in imagined locomotion. *Journal of Experimental Psychology, Human Perception and Performance, 35*, 1458-1471.
- LALLOUACHE, M.T. (1990). Un poste « visage-parole » : acquisition et traitement de contours labiaux. (*A “face-speech” workstation. Acquisition and processing of labial contours*). In *Proceedings XVIIIèmes Journées d'Études sur la Parole*, pp. 282-286.
- LIBERMAN, A.M., & WHALEN, D.H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences, 4*, 187-196.
- LINDBLOM, B., LUBKER, J., & GAY, T. (1977). Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. *The Journal of the Acoustical Society of America, 62(S1)*, S15–S15.
- MCGURK, H., & MACDONALD, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746–748.
- MEISTER, I.G. WILSON, S.M., DEBLIECK, C., WU, A.D., & IACOBONI, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology, 17*, 1692-1696.
- MÖTTÖNEN, R., & WATKINS, K.E. (2009). Motor representations of articulators contribute to categorical perception of speech sounds. *The Journal of Neuroscience, 29*, 9819-9825.
- MÖTTÖNEN, R., DUTTON, R., & WATKINS, K.E. (2013). Auditory-Motor Processing of Speech Sounds. *Cerebral Cortex, 25*, 1190-1197.

- MÜSSELER, J., & HOMMEL, B. (1997). Blindness to response-compatible stimuli. *Journal of Experimental Psychology, Human Perception and Performance*, 23, 861–872.
- MURRAY, D.J. (1968). Articulation and acoustic confusability in short term memory. *Journal of Experimental Psychology, Human Perception and Performance*, 78, 679–684.
- OPPENHEIM, G.M., & DELL, G.S. (2010). Motor movement matters: The flexible abstractness of inner speech. *Memory and Cognition*, 38, 1147-1160.
- PINTNER, R. (1913). Inner speech during silent reading. *Psychological Review*, 20, 129-153.
- REISBERG, D., SMITH, J.D., BAXTER, D.A., & SONENSHINE, M. (1989). “Enacted” auditory images are ambiguous; “pure” auditory images are not. *Quarterly Journal of Experimental Psychology*, 41, 619-641.
- ROGALSKY, C., MATCHIN, W., & HICKOK, G. (2008). Broca’s area, sentence comprehension, and working memory: an fMRI study. *Frontiers in Human Neuroscience*, 2, 1-13.
- SAMS, M., MÖTTÖNEN, R., & SIHVONEN, T. (2005). Seeing and hearing others and oneself talk. *Brain Research Cognitive Brain Research*, 23, 429-35.
- SATO, M., GRABSKI, K., GLENBERG, A., BRISEBOIS, A., BASIRAT, A., MÉNARD, L. & CATTANEO, L. (2011). Articulatory bias in speech categorization: evidence from use-induced motor plasticity. *Cortex*, 47, 1001-1003.
- SATO, M., TREMBLAY, P. & GRACCO, V. (2009). A mediating role of the premotor cortex in phoneme segmentation. *Brain and Language*, 111, 1-7.
- SATO, M., TROILLE, E., MÉNARD, L., CATHIARD, M.-A., & GRACCO, V. (2013). Silent articulation modulates auditory and audiovisual speech perception. *Experimental Brain Research*, 227, 275-288.

- SAVARIAUX, C., PERRIER, P., & ORLIAGUET, J.-P. (1995). Compensation strategies for the perturbation of the rounded vowel [u] using a lip-tube: A study of the control space in speech production. *Journal of the Acoustical Society of America*, *98*, 2428-2442.
- SCHWARTZ, J.-L., MÉNARD, L., BASIRAT, A., & SATO, M. (2012). The Perception for Action Control Theory (PACT): A perceptuo-motor theory of speech perception. *Journal of Neurolinguistics*, *25*, 336-354.
- WATKINS, K.E., STRAFELLA, A.P., & PAUS, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia*, *41*, 989-994.
- WEXLER, M., KOSSLYN, S.M., & BERTHOZ, A. (1998). Motor Processes in Mental Rotation. *Cognition*, *68*, 77-94.
- WILSON, S.M., SAYGIN, A.P., SERENO, M.I., & IACOBONI, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, *7*, 701-702.
- WOHLSCHLÄGER, A. (2000). Visual motion priming by invisible actions. *Vision Research*, *40*, 925–930.
- WOHLSCHLÄGER, A. & WOHLSCHLÄGER A. (1998). Mental and manual rotation. *Journal of Experimental Psychology, Human Perception and Performance*, *24*, 397-412.
- YETKIN, F.Z., HAMMEKE, T.A., SWANSON, S.J., MORRIS, G.L., MUELLER, W.M., MCAULIFFE, T.L., & HAUGHTON, V.M. (1995). A comparison of functional MR activation patterns during silent and audible language tasks. *American Journal of Neuroradiology*, *16*, 1087-1092.

## **Figure captions**

### **Figure 1 – Perceptuo-motor content of articulatory speech suppression**

If motor knowledge plays a role in perception (a), then a perturbing speech motor task could access motor knowledge and hence perturb speech perception (b, perturbed components in light grey); or the perturbing speech motor task could generate auditory imagery by means of efference copy and producing interference on the target stimulus to perceptually process (c, perturbed components in light grey).

### **Figure 2 – Effect of a dual-task paradigm on the generation of inner speech**

If inner speech involves the ‘motor knowledge’ network (a), an overt orofacial task mobilizing the speech production system should significantly perturb inner speech (b, perturbed components in light grey).

### **Figure 3 – Description of the experimental procedure**

Sequence of steps for each condition (see text).

### **Figure 4 – Experimental setup**

The subjects’ lips were made up with lipstick, and one small round sticker was stuck on their chin, another one on the inside of their left hand and a third one on the last joint of their middle finger. This enables to track orofacial and hand gestures over time. (The face has partially blurred for image rights reasons).

### **Figure 5 – Trajectory analysis (example provided for the lips width trajectory)**

On each trajectory it is possible to define cycles and to estimate their period and also, for lips and jaw, their amplitude. The example is provided here for lip width, in the case of

lip protrusion-retraction cycles. The period is estimated by defining a threshold (horizontal line), detecting temporal events when crossing the threshold towards larger values (black circles on the horizontal line), and computing the temporal distance between two consecutive circles. The amplitude is estimated by the maximum value inside a given cycle (displayed by filled black circles). The vertical dotted line displays the time when the subject closed the eyes and began the inner counting task.

### **Figure 6 – Results of the counting task**

Mean counting time from 1 to 30 averaged over the two counting essays and the 10 subjects, for each of the 11 conditions.

### **Figure 7 – Results of the perturbing task**

Variations of the mean and relative standard deviation of the average cycle periods for the three gestures and the two speeds (slow on the right, fast on the left) respectively before (filled black symbols) and after (black surrounded light grey symbols) eye closure. The dotted lines display the expected mean period respectively at 2 and 1 s for the slow and fast conditions.

### **Figure 8 – Compared perturbations of one task on the other**

Absolute differences between single and dual task performances for both the Target and the Perturbation tasks and for the three gesture types and the two speeds (see text). Standard deviations were computed on log values, hence they are not displayed on this figure.

### **Figure A1 – Results of the control task**

Mean time for the manual sequence (successively touching with the right hand the forehead, the chin, the left ear, the right ear and the nose, and repeating the sequence a second time in the same order) averaged over the two essays and the 12 subjects, for each of the 8 conditions.

**Table 1 – The 11 conditions of the experimental paradigm**

1	No hand/face movement	Overt counting (loudly)	
2		Covert counting	Mental counting
3	Static (no movement)	Hand	
4		Jaw	
5		Lips	
6	Dynamic (slow rate)	Hand	
7		Jaw	
8		Lips	
9	Dynamic (fast rate)	Hand	
10		Jaw	
11		Lips	