Title: Quick compensatory mechanisms for tongue posture stabilization during speech production

Abbreviated title: Tongue posture control for speech production

Authors: Takayuki Ito [1,2*], Andrew Szabados [1], Jean-Loup Caillet [1], Pascal Perrier [1]

Author's affiliations:

1: Univ. Grenoble Alpes, CNRS, Grenoble INP*, GIPSA-lab, 38000 Grenoble, France

* Institute of Engineering Univ. Grenoble Alpes.

2: Haskins Laboratories, 300 George Street, New Haven, CT, 06511, USA

*: **Corresponding author**: Takayuki Ito

Takayuki Ito

Address: GIPSA-lab, CNRS, 11 rue des Mathématiques, Grenoble Campus BP46, F-38402, Saint Martin d'Hères Cedex, France

Email: takayuki.ito@gipsa-lab.grenoble-inp.fr

Tel: +33 (0)4 76 82 70 46

Fax: +33 (0)4 76 57 47 10

Abstract (250 words)

The human tongue is atypical as a motor system since its movement is determined by deforming its soft tissues via muscles that are in large part embedded in it (muscular hydrostats). However, the neurophysiological mechanisms enabling fine tongue motor control are not well understood. We investigated sensorimotor control mechanisms of the tongue through a perturbation experiment. A mechanical perturbation was applied to the tongue during the articulation of three vowels (/i/,/e/, /ɛ/) under conditions of voicing, whispering and posturing. Tongue movements were measured at three surface locations in the sagittal plane using electromagnetic articulography. We found that the displacement induced by the external force was quickly compensated for. Individual sensors did not return to their original positions but went towards a position on the original tongue contour for that vowel. The amplitude of compensatory response at each tongue site varied systematically according to the articulatory condition. A mathematical simulation that included reflex mechanisms suggested that the observed compensatory response can be attributed to a reflex mechanism, rather than passive tissue properties. The results provide evidence for the existence of quick compensatory mechanisms in the tongue that may be dependent on tunable reflexes. The tongue posture for vowels could be regulated in relation to the shape of the tongue contour, rather than to specific positions for individual tissue points.

New & Noteworthy

This study presents evidence of quick compensatory mechanisms in tongue motor control for speech production. The tongue posture is controlled not in relation to a specific tongue position, but to the shape of the tongue contour to achieve specific speech sounds. Modulation of compensatory responses due to task demands and mathematical simulations support the idea that the quick compensatory response is driven by a reflex mechanism.

Keywords

Speech motor control, Mechanical perturbation, Reflex, Compensatory response, Orosensory feedback.

Introduction

The tongue is a fundamental organ in a variety of basic biological functions for humans, such as breathing, swallowing and speaking. Loss or damage of tongue function and its control severely affect daily life. While understanding the mechanisms of tongue control is crucial, the investigation of its sensorimotor foundations has been impeded for a long time, mostly due to the technical difficulty of having access to the tongue during its movement. Although a variety of measurement techniques (X-ray, electro-magnetometer, ultrasound devices or magnetic resonance imaging) have been proposed for the study of tongue motor control (Hoole and Zierdt 2010; Perkell et al. 1992; Westbury 1994; Whalen et al. 2005; Xing et al. 2013), investigations were limited to descriptive studies of the kinematic characteristics of tongue movements, and hence little is known about the cortical or subcortical control of these movements.

Most of the volume of the tongue is made of muscles, which are quasi-incompressible deformable bodies. As a consequence, tongue movements correspond to soft tissue deformations that are deeply constrained by this quasi-incompressibility. Thus, the tongue behaves dynamically like a muscular hydrostat (Kier and Smith 1985), which is clearly different from the behavior of a skeletal system, such as the limb or the arm, in which solid bony structures move around joints.

In case of speech production, the mechanisms of tongue motor control have been investigated especially through the adaptation paradigm using static perturbation, such as holding jaw positions with a bite block (Fowler and Turvey 1981; Gay et al. 1981), altering lip geometry with a lip tube (Savariaux et al. 1995) or modifying the palate shape by inserting an artificial palate (Brunner et al. 2011; Honda et al. 2002; McFarland et al.

1996). The produced sounds become quite similar to the original sounds after a short time of adaptation following the perturbation, suggesting that the tongue control system quickly adapts to new environments as we experience in daily life situations (e.g. speaking while chewing gum). While those adaptive changes provide evidence for efficient and robust control of the tongue, little is known about the neurophysiological and physical phenomena enabling the control of tongue posture stabilization during the production of speech sounds, in particular to compensate for time-varying perturbations as experienced in our daily life.

In regard to speech production, a well-known hypothesis suggests that tongue motor control consists in moving the tongue toward target postures associated with phonemes (Keating 1990; MacNeilage 1970; Patri et al. 2015). In the production of consonants, such as /t/ or /s/, the tongue must not only target postures precisely, but also maintain them for a sufficient amount of time, which requires the capacity to efficiently stabilize the tongue in the target postures. This task could be accomplished by moving the tongue, or a part of the tongue, toward a specific location in the oral cavity, such as toward the upper teeth, just as it is done in pointing or reaching tasks with the arm by moving the different arm segments around their joints. It is however unclear whether the postural stabilization of the deformable tongue relies on control principles that are similar to those of the stabilization of arm segments around a joint.

The control mechanisms involved in limb and body posture stabilization have often been investigated using time-varying perturbations (Burdet et al. 2001; Gomi and Osu 1998; Gottlieb and Agarwal 1988; Marsden et al. 1972; Pruszynski and Scott 2012; Soechting 1988; Soechting and Lacquaniti 1988). Immediate compensation for time-

varying perturbations can be driven by impedance control and neural feedback control. Responses associated with impedance control are intrinsic physical responses of the motor plant associated with passive elastic properties (Gomi and Osu 1998; Gottlieb and Agarwal 1988). As a consequence, they are very fast. The response induced by neural feedback control, such as stretch reflex, arrives later due to the delay in neural transmission, which ranges from 20 to 50 ms for short-latency stretch reflex and from 50 to 100 ms for the long-latency stretch reflex (Pruszynski and Scott 2012). This time sequence of responses was examined in the speech articulatory system using a mechanical perturbation of the jaw (Gomi et al. 2002; Ito et al. 2005). The compensatory response due to stiffness in the linkage between lip and jaw occurred less than 20 ms after the onset of the movement induced by the perturbation; in contrast, the muscle activation due to reflex was induced longer than 50 ms after this movement onset (Gomi et al. 2002). This compensatory reflex was mediated by the motor cortex (Ito et al. 2005). As in limb studies, applying a time-varying mechanical perturbation to the tongue and observing the behavioral response is an efficient approach to examine whether the tongue shows compensatory responses with latencies comparable to those of responses induced by impedance control and by neural feedback control in the limb system.

Task-dependent modulation can be seen in both impedance control and reflex mechanisms. In impedance control, mechanical properties of the motor system (such as stiffness or damping) can be adjusted for external constraints (Burdet et al. 2001; Gomi and Osu 1998). It has been found in the speech articulatory system that the stiffness in the linkage between the jaw and upper lip increased for the production of the bilabial fricative consonant /Φ/ as compared to the neighboring vowels in vowel-consonant-vowel

129 sequences (Gomi et al. 2002). In reflex control, long latency reflexes in the limb system
130 have been shown to vary depending on the task demands in a variety of situations (See
131 review in Pruszynski and Scott (2012)). In the orofacial system, the magnitude of
132 compensatory reflexes in the upper lip has been shown to vary depending on the speech
133 task: the magnitude of the muscle activation due to reflex was significantly larger during
134 the production of bilabial fricatives than during vowel production (Gomi et al. 2002), and
135 little or no response was induced in a resting condition (Ito and Gomi 2007). Similarly,
136 Kelso et al. (1984) observed an increase of muscle activation in the tongue in response to
137 a force perturbation applied to the jaw during the production of alveolar consonants /z/,
138 but not during the production of bilabial consonants /b/. Thus, compensatory responses
139 induced in the tongue by a force perturbation applied to the jaw seem also to be
140 modulated across different speech tasks and/or across speech and non-speech (resting)
141 tasks.

142 In speech production, a number of converging experimental and modeling results
143 suggest that tongue stiffness varies across the front vowels, such as /i/, /e/ and /ɛ/, in
144 conjunction with the amount of tongue positioning variability: the stronger the stiffness,
145 the smaller the variability. Previous studies (Beckman et al. 1995; Perkell and Nelson
146 1985) showed that front-high vowels /i/ and /e/ have less articulatory variability than
147 lower vowels like /ɛ/ and /a/. In addition, using electropalatography (Stone and Lundberg
148 1996) and MR imaging (Badin et al. 2002), a strong relationship was found between the
149 intensity of the linguo-palatal contact and the activation of the posterior genioglossus,
150 which pulls forward the back part of the tongue, as evidenced by magnitude of grooving
151 the back part of the tongue. The idea is also supported by the simulations with a 3D

biomechanical model of the tongue showing that, for a given height of the tongue, an increase in the amount of lateral contacts between the tongue and the palate is achieved through an increase in tongue muscle activation, specifically activation of the posterior genioglossus (Buchaillard et al. 2009; Gick et al. 2017). An increase in muscle activation induces an increase in muscle tissue stiffness called the stress-stiffening effect (Nazari et al. 2011). Applying a mechanical force perturbation to the motor system, as used in previous studies (Burdet et al. 2001; Gomi and Osu 1998; Gottlieb and Agarwal 1988), allows stiffness to be measured by observing the displacement induced by the perturbation, since for a given magnitude of the force the magnitude of the displacement is proportional to the stiffness, assuming linear dynamical properties. Hence, force perturbation can be an efficient tool to reveal task-dependent changes in stiffness of the tongue or of the linguo-palatal system.

For the reflex mechanism, it is still unclear whether or not a stretch reflex or similar types of reflex are involved in the postural control of the tongue. In the tongue, it is still under debate which sensory receptors play a role in proprioception and in stretch reflex. In general, muscle spindles are considered as the main source of proprioception and stretch reflex (Schmidt and Lee 2005). Cooper (1953) had found muscle spindles in the muscles of the human tongue. However, Neilson et al. (1979) did not find any evidence that tonic stretch reflex was induced in the tongue in response to stretching of the tip of the tongue. In this context, we can assume that muscle spindles in the tongue may not work like those in the skeletal muscles of the limbs. Showing a compensatory response with an appropriate latency can provide a clue for the involvement of reflex

mechanism in the control of posture stabilization and can suggest how muscle spindles or proprioceptive organs are involved in the compensatory responses.

In our study, we focus on quick compensatory responses for the stabilization of the tongue. We examined whether perturbations of tongue postures were compensated by on-line feedback control mechanisms by observing the time course of the tongue response due to time-varying perturbations. For this purpose, a novel mechanical perturbation system was specifically developed using a robotic device. We applied the mechanical perturbation in an unanticipated manner while subjects were holding specified tongue postures. Electromagnetic articulography was used to measure the displacement of the articulatory organs (jaw, lips and tongue).

We expected to observe different phases over time in the compensatory responses to the perturbation, with different latencies, due to the successive involvement of different feedback mechanisms: impedance control associated with mechanical stiffness of the tongue muscles, short delay stretch-like reflexes, and long delay auditory and/or somatosensory feedback. We also expected to observe a variation in the magnitude of the displacement induced by the perturbation according to (1) the vowel posture, because of differences across vowels in the stiffness of tongue muscles, and (2) articulatory manners (speaking versus non-speaking), because of differences in the requirements in terms of postural control accuracy between these two tasks.

In order to analyze the time-course of the responses to perturbation, especially in terms of the shape of the kinematic responses respectively induced by impedance control and by stretch-like reflex, we simulated the tongue force perturbations using a mathematical biomechanical model of the tongue (Perrier et al. 2003). Since this

mathematical simulation includes an account of short-delay somatosensory feedback for reflex response, it can be used to dissociate a reflex contribution from the response by impedance control through manipulation of the amplitude of the reflex gain in the model. By comparing the simulated movements of the tongue in the model with those observed experimentally, we could assess whether the reflex mechanism as implemented in the model could explain the experimentally observed compensatory responses.

Methods

Participants

Nine native French speakers (18-35 years old, 3 females) participated in the experiment. The participants were all healthy young adults who reported normal hearing. All participants signed informed consent forms approved by the local ethical committee of the University Grenoble Alpes (CERNI: Comité d'Ethique pour les Recherches non Interventionnelles: CERNI-AvisConsultatif-2017-01-17-04).

Articulatory measurement and tongue perturbation

Electromagnetic articulography (Wave, Northern Digital Inc.) was used to measure articulation. For vowel production, the articulation is primarily characterized by the geometry of the vocal tract shape in the mid-sagittal plane. Hence, the measurement sensors were glued in the mid-sagittal plane: tongue tip (TT), tongue blade (TB), and tongue dorsum (TD) on the tongue surface, upper and lower lips (UL and LL), and jaw (J). Four reference sensors, located on the nasion, the left and right mastoids and the upper incisors (see Fig. 1), were also used for the measurement of head movements and

their subtraction from the other sensors' movements in an off-line analysis. The sensor data were recorded with a 400 Hz sampling rate, and the produced speech sound was recorded synchronously with a 22.05 kHz sampling frequency.

A small robotic device (Phantom Premium 1.0, Geomagic) was used to apply a force to the tongue (see Figure 1A). The robot was set in front of the subject and connected to the tongue surface through a thin thread. At both extremities of the thread, two small anchors (see TP in Figure 1B) were glued on the tongue surface on both sides of the tongue blade sensor (TB) of the articulograph. The tongue perturbation was produced by pulling the tongue forward with a 1 N force for 1 second. The force was applied with 5-ms rising and falling phases in order to avoid mechanical noise of the robot. The subject's head was held in place by a head holder in order to maintain a constant stable position.

Experimental tasks: vowels and articulatory manners

The first focus of this study is to compare perturbed responses in different vowel postures. We focused on three French vowels, /i/, /e/ and /ɛ/ that are expected to be produced with different amounts of tongue muscle activation. Vowels /i/, /e/ and /ɛ/ are close neighbors in terms of both acoustical and articulatory characteristics. In the acoustical domain, vowels are efficiently described by the first two formants, F1 and F2, which are the lowest two resonance frequencies of the vocal tract and correspond to the first two maxima of the frequency spectrum of the sound. The first formant gradually decreases from /ɛ/, to /e/ and /i/, while the second formant increases. There is no other vowel between them in French phonetics. In the articulatory domain, these vowels are

usually discriminated along the tongue height (or jaw height) dimension, which corresponds to the vocal tract aperture. The production of /i/ requires a high and anterior tongue position, corresponding to the narrowest aperture of the vocal tract in the palatal region. As compared to the production of /i/, tongue height gradually decreases and moves backwards for the production of /e/ and /ɛ/. This variation is associated especially with a decrease in the activation of the posterior genioglossus.

The second focus is to examine the effect of speaking versus non-speaking conditions. We compared compensatory responses in voicing, whispering and posturing conditions. In the three conditions, the subjects were asked to sustain the tongue position for a few seconds. In voicing and whispering conditions, they were asked to do so while voicing or whispering. In non-speech posturing condition, they were asked to do so silently. In order to have the same set of reference vowel tongue positions as for the speaking conditions, subjects were instructed to produce the vowel briefly at the beginning of each trial of the posturing condition, and then to maintain the posture silently (without voicing or whispering).

Experimental procedure

Nine experimental tasks (three vowels under three articulatory manners) were tested in random order. At the beginning of each trial, visual instruction on a monitor was used to inform the subjects about the task for the coming trial. The trial was launched by the experimenter after examining whether the participant was ready for the task. The subjects were instructed to start their task via visual and auditory cues. The perturbation was applied 1s after the trial onset on a randomly-selected third of the trials in order to

avoid any anticipation by the subject. There were 270 trials in total. As a result, 10 perturbed responses were recorded in each condition. The total experimental time was approx. 2 hours including the preparation time for placing the EMA sensors and gluing the wires of the robotic arm on the participant's tongue, and the experimental session itself. The long duration of the experiment and the complexity of the tasks made it difficult to record a larger number of subjects.

Data analysis

For pre-processing of articulatory movement data, three-dimensional head movement correction was done off-line based on the head position data provided by reference sensors. We also subtracted the movement information of the jaw from the tongue and lower lip values since those are carried by the jaw. Velocity and acceleration were calculated to detect changes in movement. As seen in Figure 2 and beyond, the forward direction of movement corresponds to a decrease of the measured values in the presented figures. The failed trials were excluded from the analysis (no exclusion for 1 subject, one exclusion for 3 subjects, two exclusions for 5 subjects).

For the data analysis, movement data were averaged across subjects. In this averaging, we did not consider raw position data due to large inter-individual differences in size and shape of the tongue and the vocal tract. Instead we took the average of the displacement relative to the position of the tongue at the time of the perturbation onset, which was calculated by aligning the data at this time to zero. We also estimated the tongue contour for each produced tongue posture by calculating direction vectors between two neighbor sensors (TB and TT; TB and TD), although this is a rough

289 estimation since we have only three sensors. The angles of those vectors were also
290 averaged across subjects.

291 In the temporal patterns of the perturbation response, we expected to observe
292 multiple phases due to the combination of reactions associated with purely passive and
293 reflex components of the tongue. To assess this multiple-phase organization of the
294 responses, we divided the perturbation response into four intervals based on evident local
295 peaks on the horizontal displacement of the TB sensor. Those intervals are shown in
296 Figure 2, numbered from R1 to R4 (See Results for detailed explanation of the extraction
297 of the intervals). The locations of these local peaks were extracted at the point when the
298 absolute value of horizontal velocity became sufficiently small (less than 10 mm/s). In
299 our analysis we mostly focused on intervals R1, R2 and R4 since those intervals reflect
300 response mechanisms that were at the core of this study: response due to passive
301 components and task-related compensatory response. The possibility of the existence of
302 overlapping responses with different latencies was further examined by looking for any
303 inflection point in the compensatory response. We then looked for peaks in the absolute
304 value of the horizontal acceleration to detect subtle changes in movement characteristics.
305 For statistical analysis, the amplitudes of tongue displacement in the sagittal plane within
306 each of the intervals were quantified using both vertical and horizontal displacements,
307 based on the detected peak points in each response phase. A mixed-effect model was
308 applied with three fixed effects: vowel (/i/, /e/ and /ɛ/), articulatory manner (voicing,
309 whispering, and posturing) and sensor-location (TT, TB, and TD). Individual variability
310 across subjects was taken into consideration as a random effect. Post-hoc tests were
311 carried out with Bonferroni correction ($p < 0.05$).

For the acoustical analysis, we extracted the first three formants (F1, F2 and F3) using Linear Predictive Coding analysis (Rabiner and Schafer 1978). Although vowels /i/, /e/ and /ɛ/ can be efficiently separated in the (F1, F2) plane, we included F3 because it provides informative cues about local articulatory changes around these vowels. This analysis was only used for the data recorded under the voicing condition. For statistical analysis, the amplitudes of F1, F2 and F3 were calculated at the representative time points that were defined on the basis of the tongue displacement peaks mentioned above. We examined whether the sound produced at these key time points of the response to the perturbation was significantly different from the sound just before the perturbation onset ("original sound" henceforth). We carried out a paired t-test between the sound characteristics at the key time points and those of the original sound.

Simulations with the 2D biomechanical model

We carried out mathematical simulations using a 2D biomechanical model of the tongue. The model was originally developed in previous studies (Payan and Perrier 1997; Perrier et al. 2003), and implements both passive stiffness and a stretch-like reflex to produce accounts of tongue kinematics in speech production. Details are described in those previous studies. Briefly, the tongue in the mid-sagittal plane is modeled as a biomechanical finite element model with 221 nodes (See Fig. 1C). Seven primary intrinsic and extrinsic muscles of the human tongue (posterior and anterior genioglossus, hyoglossus, styloglossus, inferior longitudinalis, superior longitudinalis, and verticalis) were embedded in the model as force actuators acting via macrofibers on selected elements, whose stiffness increases with muscle activation. Details on the insertion points

335 and muscle description can be found in Perrier et al. (2003). As representative examples,
336 we show the implementation of the styloglossus and posterior genioglossus (Figure 1C)
337 since these two muscles are considered to have a major role in the response to the
338 mechanical perturbation used in this study. The styloglossus is the primary muscle for
339 tongue retraction and also for the reaction against our study's perturbation, which was
340 applied in the direction of tongue protrusion. The posterior genioglossus is the main
341 muscle responsible for the tongue elevation and protrusion associated with the production
342 of the front /i/, /e/ and /ɛ/ vowels.

343 The upper contour of the tongue mesh model has 16 nodes, which are numbered
344 in increasing order from the tongue root to the tongue tip. Nodes 11, 13 and 16 were
345 chosen as the corresponding locations of the sensors used in the test, TD, TB and TT,
346 respectively (See Figure 1C). The perturbation force was applied to four nodes consisting
347 of nodes 13 and 14 and two nodes below those (TP in Figure 1C) in the horizontal
348 direction. Due to numerical instability issues in the numerical simulation and to
349 differences between the experimental tongue configurations and the modeled ones, we
350 applied a perturbation force of 0.25 N in each perturbed node for a shorter duration (200
351 ms) for the production of vowel /e/.

352 In the model, muscle force is generated based on the λ-model (Feldman 1986) as
353 follows:

354 $$F(t) = \rho(e^{cA} - 1)$$

355 where A is the muscle activation level, c is the reflex gain parameter, and $\rho(\cdot)$ is
356 the factor accounting for the force generation capacity of each muscle, which is a
357 function of its cross-sectional area and its fiber density in each muscle. The muscle

activation level A is itself determined as a function of the current muscle length L and a value λ', which depends on the threshold muscle length λ for activation and on the rate of change in muscle length ($\lambda' = \lambda$ then in static configurations), as follows:

$$A = \begin{cases} L - \lambda', & L > \lambda' \\ 0, & L \leq \lambda' \end{cases}$$

To evaluate the potential role of the embedded short-latency somatosensory reflex in the tongue response to the perturbation, we carried out a series of simulations in which we modified the amplitude of this reflex. The simulation always started from the posture at rest (zero muscle forces), which corresponds to the posture for the production of a schwa. In the formula shown above, the muscle activation level includes both descending activations adapted to the achievement of the posture for a specific vowel and reflex activation. In order to control the gain for reflex activation separately, we reformulated muscle activation A in two parts as follows:

$$F(t) = \rho\left(e^{c(L_e-\lambda)+kc(L-L_e)} - 1\right)$$

where L_e and λ are the actual length and the centrally specified motor command of the muscle in the tongue posture selected for vowel /e/. These L_e and λ values were determined in previous studies (e.g. Perrier et al., 2003). The activation $A_e = L_e - \lambda$, corresponds to the baseline activation level, reached in vowel /e/ before the perturbation started. The reflex activation induced by the perturbation was represented as the difference between the muscle length L measured during the response to the perturbation and the baseline muscle length L_e. In this formula, k enables controlling the gain of the reflex activation induced by the perturbation. We tested in the simulations several amplifications of this gain, from 0 to 10. Condition $k = 0$ simulates the situation in

which reactions to the perturbation are strictly limited to the consequences of passive mechanical components.

Results

Overview of the time course of the response in the direction of the perturbation

We first considered the time course of the response to the perturbation in the direction of the perturbation, i.e. along the horizontal axis. Figure 2A shows the averaged horizontal displacement for vowel /ɛ/ under voicing condition for all the sensors (UL, LL, J, TT, TB and TD). Increasing values correspond to a backward horizontal movement. The vertical dashed lines indicate the displacement peaks delimiting intervals R1 to R4 on the TB sensor, and the dotted line marks for the same sensor an inflection point that is interpreted as evidence for an overlap of responses with different latencies. As shown here, the perturbation induced a large tongue displacement in the forward direction (R1) and a certain amount of this change was then recovered by a compensatory movement (R2). After reaching the peak compensatory response on TB (end of R2), the tongue could not precisely hold the same posture; rather, the tongue gradually drifted forward (R3). This indicates that the tongue posture was more efficiently controlled to resist against the transient change, than against the constant bias force. This may be due to the fact that a constant force constraint is quite rare in daily life, whereas transient changes are often experienced when biting and drinking. As soon as the perturbation force was removed, the tongue moved in the backward direction, but it did not completely return to its original position (R4). This aspect of the response amplitude will be discussed later in the section "*Geometrical path of the response in the mid-sagittal plane"*. The upper and

lower lips, and the jaw also showed significant displacements in the R2 interval, although the amplitude was considerably smaller than the ones in the tongue. Considering that there is some delay between the perturbation onset and the displacement onsets of these articulators, these displacements may be interpreted as a consequence of a heterogenic reflex loop arising from muscles that are not in the tongue.

Figure 2B represents a magnified view of the compensatory response in TB as representative data. This plotted part of the signal began 50ms after the perturbation onset and lasted 300ms, as marked by the black thick horizontal bar on the x-axis in Figure 2B. During this time interval, there was an inflection point associated with a peak in the second derivative at around 140ms after the perturbation onset (top panel in Figure 2B, vertical thin-dotted line), indicating a significant change in the response at this time. We also observe an interesting velocity profile (middle panel in Figure 2B) with two peaks of velocity, one before and one after the inflection point, indicating that the compensatory response was not a single phase response, but rather a sequence of at least two responses: The first component may be a passive reaction due to the passive stiffness of the tongue, and the second component may result from an active reflex-based compensatory response with some delays due to neural transmission. Based on this observation, we divided the compensatory response in two parts, before (Interval R2-1) and after (Interval R2-2) the inflection points on the displacement curve. We analyzed these two compensatory responses separately.

Geometrical path of the response in the mid-sagittal plane

Next, we considered the displacement of the tongue in the mid-sagittal plane. Figure 3A represents the averaged trajectory of sensor TB. The three panels represent the three vowel conditions. In each panel, the different lines represent the different articulatory manners (voicing, whispering and posturing). We aligned the data at zero at the onset-time of the tongue perturbation. The displacement peaks, which are represented by the vertical dashed lines in Figure 2, are presented by asterisks in Figure 3A.

In general, the detailed responses from the onset to the offset of the perturbation are quite similar across articulatory manners and vowels. We can also confirm the temporal structure of the response observed in the previous section as follows. Once the perturbation was applied in the forward direction, the tongue moved first in the same direction (initial response, interval R1). The precise direction of this initial response varied slightly depending on the articulatory configuration at the time of the perturbation onset, because of slight differences in the orientation of the wire connected to the robotic arm due to differences in jaw opening. After reaching the peak displacement of the initial response, the tongue moved back in a compensatory response (R2-1 and R2-2). Then, the position gradually drifted mostly in the direction of the perturbation force (R3). When the perturbation force was removed, the tongue moved back in the posterior direction (R4).

The trajectories in the mid-sagittal plane revealed that the compensatory response in the interval R2 did not tend to bring the sensor back to its original position, because of a significant and continuous lowering of the tongue. To analyze this diagonal direction of the compensatory movement, we superimposed our rough estimation of the original tongue contour at the perturbation onset on the sensor trajectories in Figure 3A (light-dashed lines in each panel). Based on this original contour, we found that the tongue

448 tended to return to its original contour, instead of its original position. The direction of
449 this compensatory movement was nearly perpendicular to the original tongue contour
450 (81.7 degrees ± 2.02). This tendency was consistent across all articulatory manners and
451 all vowels. This suggests that the compensatory mechanism underlying tongue posture
452 control is not to maintain or recover the exact whole same position of the tongue, but
453 rather to maintain the same local shape of the tongue contour in the region described by
454 the sensors. This statement is also supported by the observation of the tongue position
455 after the perturbation removal, since the final positions of the sensors did not match the
456 original positions, but were close to the original tongue contour, specifically in the
457 speaking conditions (voicing and whispering).
458
459 ***Quantitative analysis of the response amplitudes***
460 Displacement amplitudes in each response phase were statistically compared
461 across tasks using linear mixed-effect models with three fixed effects [vowel (/i/, /e/ and
462 /ɛ/), articulatory manner (voicing, whispering, and posturing) and sensor location (TT,
463 TB and TD)]. We found an interaction between sensor location and articulatory manner
464 in the later part of the compensatory response (R2-2) [$F(4,182) = 7.32$, $p < 0.001$]. All
465 other contrasts (R1, R2-1, R3 and R4) did not show any reliable interaction effect.
466 Accordingly, we present first the compensatory response in the interval R2-2, and then
467 the fixed effects in other contrasts, particularly in R1, R2-1 and R4.
468 For a better understanding of the interaction effect in R2-2, we took an average
469 across the vowels as shown in Figure 4. The pattern of the amplitude variation across the
470 sensor locations was different in each articulatory manner. In the voicing condition, the

amplitudes of tongue tip (TT) and blade (TB) movements were larger than that of the tongue dorsum (TD) ($p < 0.001$ in both), and TT and TB were not different from each other ($p > 0.2$). A similar tendency was found in whispering condition ($p < 0.001$ between TT and TD and between TB and TD), but TT and TB were modestly different ($p = 0.08$). On the other hand, in the posturing condition, there was no significant difference between TT and TD and between TB and TD ($p > 0.9$), but there was a significant difference between TT and TB ($p < 0.05$). When we compared the differences across articulatory manners in each tongue location, TT showed larger amplitude of movements for compensation when subjects were producing speech sounds ($p < 0.001$). Contrarily, TD moved less when subjects were speaking ($p < 0.05$). TB was not significantly different across the articulatory manners ($p > 0.5$). The tendency of reduced compensatory response in TD was more remarkable along the vertical direction ($p < 0.001$ both for voicing *versus* posturing and for whispering *versus* posturing), but not along the horizontal one ($p > 0.4$ in all three comparisons). This suggests that the posterior section of the tongue may be controlled to maintain an appropriate amount of constriction opening between tongue and palate for speaking, which would induce a small tolerance to changes in response to the perturbation, unlike what is observed for TT. In summary, our results suggest that the gain of the compensatory response was specifically increased in the front part of the tongue (TT) and decreased in the back part of the tongue (TD) in both speaking conditions. This is consistent with the idea that the tongue shape is controlled in a systematic way to preserve tongue characteristics that are crucial for targeted speech sounds.

For the fixed effect components in all contrasts, we took an average in each fixed effect as shown in Figure 3B. We found a reliable difference across the different levels of the fixed effects (vowel, articulatory manner and sensor location). As an overall tendency, the influence of each fixed effect on the amplitude of tongue displacement was similar in the intervals R1, R2-1, and R4, and the R2-2 intervals were different from the others. This also supports the idea that the compensatory response in R2-2 may be driven differently from the ones in R1, R2-1, and R4.

In regard to the vowel influence (top panel in Figure 3B), the displacement amplitude was reliably different ([$F(2,182) = 175.5$, $p < 0.001$] in R1, [$F(2,182) = 13.45$, $p < 0.001$] in R2-1 and [$F(2,182) = 142.2$, $p < 0.001$] in R4). The largest displacement was seen in the production of /ɛ/ and the amplitude of displacement was gradually reduced for /e/ and /i/. The post-hoc test showed there was a significant difference in all of the pairwise comparisons ($p < 0.02$) consistently in the three response intervals R1, R2-1 and R4. In the comparison of the three articulatory manners we observe that the displacements in whispering and voicing conditions were consistently smaller than those in the posturing condition in the initial part of the perturbation response (interval R1) [$F(2,182) = 24.01$, $p < 0.001$]. There was no reliable difference between voicing and whispering ($p > 0.9$). A similar tendency in the difference between the voicing and whispering manners and the posturing manner was also found in the earlier phase of compensatory movement [$F(2,182) = 4.823$, $p < 0.01$]. In R4, the averaged values showed a similar tendency although no difference was found across the three articulatory manners [R4: $F(2,182) = 1.774$, $p > 0.15$].

Finally, we found a significant difference according to the sensor location on the tongue [$F(2,182) = 26.22$, $p < 0.001$] in R1 (bottom panel in Fig. 3B). The displacement in the tongue blade (TB) was not different from the one in the tongue dorsum (TD) ($p > 0.9$), whereas the displacement in tongue tip (TT) was smaller than the other two ($p < 0.001$ in both). This was also seen when the perturbation was removed [R4: $F(2,182) = 22.74$, $p < 0.001$], and a modest difference was found in R2-1, [$F(2,182) = 2.384$, $p = 0.09$]. The small movement amplitude of the TT sensor does not seem to be due to contact with the teeth, since it was observed both in the forward and backward direction after the onset and the offset of the perturbation.

Formant patterns in speech sounds

The changes in tongue shape due to the tongue perturbation also resulted in changes in the produced sounds. Particularly, a compensation for the perturbation in the space of the first three formants was observed in relation with the amount of compensatory movement. Figure 5 represents the temporal pattern of the three formants averaged across the participants. Triangles represent the times at the key points in the horizontal movement as mentioned above (dashed and dotted lines on Figure 2), with the first triangle corresponding to the end of interval R1 and the beginning of interval R2. The acoustic impact of the perturbation varied across vowels. The peak changes occurred at the onset of the reflex-like compensatory movement (second triangle corresponding to the beginning of interval R2-2). The largest changes were induced during the production of the vowel /ɛ/.

For vowel /ɛ/, the changes at the first negative peak were significant for F1 and F3 ($p < 0.005$ for both), but not in F2 ($p > 0.3$). Those were significantly reduced at the peak displacement of the compensatory response (third triangle corresponding to the end of interval R2-2; $p < 0.03$ for F1 and $p < 0.002$ for F3), although their frequencies were still significantly different from their values at the perturbation onset ($p < 0.003$ for F1 and $p < 0.05$ for F3). When the perturbation was removed, the changes between the fourth and fifth triangles were induced again in F1 and F3 ($p < 0.03$ for F1 and $p< 0.002$ for F3). For vowel /e/, the initial change due to the perturbation was induced in all three formants (F1: $p< 0.05$, F2: $p<0.002$, F3: $p< 0.005$). At the end of the compensatory response R2 (third triangle), the changes in F1 and F3 were significantly reduced (F1: $p< 0.01$, F3: $p<0.0001$) with no reliable difference from their original frequencies ($p > 0.1$ for F1 and $p > 0.4$ for F3). On the contrary, the compensatory movement did not reduce the alteration of the second formant induced by the perturbation ($p > 0.6$), which remained even after the removal of the perturbation. The removal of the perturbation changed F1 ($p < 0.01$) and F3 ($p < 0.001$), but not F2 ($p > 0.2$). For vowel /i/, the initial change was induced in F1 and F2, but not in F3. Those formants returned to their original values at the perturbation onset by the compensatory response.

Interestingly, the early component R2-1 of the compensatory response (between the first and second triangle) was not effective in compensating the acoustic effects of the perturbation. Rather acoustic compensation occurred mostly in the second component R2-2 of the compensatory response (between the second and the third triangle). This suggests the second component of the compensatory response (R2-2) may be specifically controlled to maintain and recover the intended speech sound.

Taken together with the displacement results, these results suggest that the tongue position is controlled not to maintain a specific tongue position, but to maintain the shape of the tongue surface in the region that is crucial for the perceptual characteristics of the sound.

Simulation with the biomechanical model: effect of the reflex gain

We carried out simulations of the force perturbation of the tongue using the 2D biomechanical model and examined the model's perturbation response. The general pattern of movement along the horizontal direction was very similar to the experimental one (see Figure 6A, $k = 1$). The trajectory showed a first immediate forward movement (interval R1) followed by a backward movement (interval R2) and a strong final backward movement once the perturbation was removed (interval R4). However, the gradual drift experimentally observed during the holding phase in interval R3 did not exist in the simulation: the tongue model remained stable.

We then tested how the compensatory response changed depending on changes in reflex gain. Figure 6A shows the temporal pattern of horizontal displacement in simulation with several reflex gains ranging from 0 to 10 times the gain in the original tongue model. Change in reflex gain affected the responses to the perturbation. The strongest consequence was observed in the amplitude of the compensatory response corresponding to interval R2 in the experimental data. The largest response was obtained with the largest reflex gain ($k = 10$). On the other hand, when the gain was set to zero, which corresponds to the situation where the tongue responded with its passive stiffness only, the compensatory response was not observed. This finding supports the idea that the

583 compensatory response is not driven by a passive component alone, but by an additional
584 neural feedback mechanism, possibly by a reflex arc. Figure 6D displays the
585 compensatory amplitudes in each observed location on the tongue contour. The amplitude
586 of TT was smaller than the other two, TB and TD and the amplitudes of TB and TD were
587 almost the same. This simulated variation across the locations on the tongue was similar
588 to our behavioral observations under posturing condition (right panel in Fig. 4). In the
589 current simulations, the reflex gain was set to the same value in all the tongue muscles.
590 The differences between the results of our simulations and the experimental observations
591 in speaking condition suggest that our subjects could have regulated the reflex gain
592 specifically in each muscle to achieve the stability requirement of the speech production
593 tasks.

594 The simulated muscle activation patterns are shown in Figure 7. The model has
595 seven tongue muscles (GGp: Posterior genioglossus, GGa: Anterior genioglossus, Hyo:
596 Hyoglossus, Sty: Styloglossus, SupLongi: Superior longitudinalis, IntLongi: Inferior
597 longitutinalis, and Vert: Verticalis). Figure 7A represents the temporal patterns of the
598 changes in muscle activations after the perturbation. The values represent relative
599 changes from the baseline activations in the tongue posture at the perturbation onset with
600 various amplitudes of the reflex gain. The top panel is a magnified view of the
601 corresponding tongue blade (TB) displacements that is already shown in Figure 6A.
602 Figure 7B showed the baseline activation levels in tongue posture of vowel /e/ reached at
603 the perturbation onset.

604 In these four muscles, we found that the activation of the styloglossus and the
605 posterior genioglossus was most affected by the perturbation, while the activation of the

606 anterior genioglossus and the inferior longitudinalis changes relatively little. The
607 styloglossus muscle showed excitatory patterns of activation, while the posterior
608 genioglossus showed inhibitory pattern of activation in contrast. These observations are
609 consistent with the fact that the forward displacement of the tongue due to the
610 perturbation caused an elongation of the fibers of the styloglossus, which increases the
611 difference between the actual fiber length and the corresponding threshold muscle
612 command λ, and a decrease in the length of the fibers of the posterior genioglossus,
613 which reduces the difference between the actual fiber length and the corresponding λ
614 command. As a result, styloglossus activation increases and pulls the tongue backward in
615 order to change the fiber length back to a value closer to the λ command, and the
616 activation of the posterior genioglossus is reduced. Although there were reflex activations
617 in the hyoglossus and verticalis, this effect is small because, due to their orientation in the
618 tongue, their fibers were less impacted by the forward displacement of the tongue.

619 We also found differences between recorded and simulated responses in the mid-
620 sagittal plane. Figure 6B shows the simulated trajectory in the sagittal plane. As in the
621 recorded data, the tongue was moved horizontally by the perturbation force. Then the
622 compensatory response was induced in the direction of the original position. After the
623 release of the perturbation force, the tongue almost returned to the original location.
624 Contrary to our experimental observations, the simulated compensatory response did not
625 induce the tongue to recover its original contour, but rather to return directly to its
626 original shape at the perturbation onset.

627 We examined the amplitude of the initial displacement immediately after the
628 perturbation (interval R1 in the experimental data) at three tongue sites (TT, TB and TD)

(Figure 6C). The largest amplitude was seen in TB, where the perturbation was applied. The displacements in TT and TD were almost the same and they were smaller than in TB. These differences could originate in the fact that the tongue model includes an isotropic account of tongue biomechanics (see Nazari et al. 2013, for a more realistic account).

Discussion

The current study enabled the first observation of a rapid compensatory response in the tongue induced by an external mechanical perturbation. The mathematical stimulations using a 2D biomechanical model of the tongue strongly support the hypothesis of the existence of a reflex mechanism by showing that a sequence of compensatory responses similar to the one observed experimentally can be reproduced with a reflex arc. In addition, the reflex-like part of the simulated compensatory response was systematically modified in vowel posturing depending on the task. More specifically, the anterior part of the tongue showed greater amplitudes of the compensatory response in the speaking tasks than in the posturing task. In contrast, the posterior part of the tongue showed a smaller response in the speaking tasks. Speaking requires more precise control to maintain the target tongue contour. Our experimental observations also suggest that the control of the tongue for vowel production does not aim at producing a specific tongue position, but rather at preserving the shape of tongue contour in the region close to the palate. These compensatory movements changed the formant patterns of the perturbed speech signal and enable recovering formant patterns that were closer to the original ones. As an additional finding, the current study quantitatively demonstrated the empirically known fact that tongue stiffness can be changed according to the vowel and

652 the articulatory manner, and that it influences the stability of the tongue posture. The
653 limited number of subjects (nine) included in this experiment obviously calls for some
654 caution in relation to these general conclusions. However, the consistency of the response
655 across all the participants and the systematic nature of the observations across repetitions
656 argue strongly for the generalization of our results.

657 Tongue motor control mechanisms have been investigated mostly based on
658 empirical observations of tongue kinematics and inspired by findings from other organs
659 such as the limb. Applying a newly developed mechanical perturbation to the tongue
660 during postural control aims at allowing the discovery of detailed clues about tongue
661 control mechanisms, such as error correction for movement and posture stabilization by
662 feedback loops, and update of feedforward commands by adaptation and learning. In the
663 current study, this approach enabled us to show, for the first time, a rapid compensatory
664 response with a strong evidence for the role of a neural feedback loop. Interestingly in
665 our experiment, this feedback control does not seem to preserve the exact tongue posture,
666 but to preserve the shape of the tongue contour, which is crucial for vowel production.
667 Such a feedback mechanism is difficult to predict from the generalization of control
668 mechanisms observed in skeletal systems, in which restoring the original position is the
669 basis of postural control. This flexibility (or redundancy) may be a special feature of
670 tongue posture control due to the muscular hydrostat system. As a possible mechanism, a
671 basic muscle tonus of the genioglossus, which is the largest muscle in the tongue
672 musculature (Takemoto 2001), may be involved in the stabilization of the original tongue
673 contour. Due to the fan-like orientation of the fibers of the genioglossus, this basic tonus
674 applies a main stress that acts in the direction to the tongue root for maintaining the

675 tongue posture in usual situations and this may largely contribute in the studied
676 compensatory movement to maintain the same tongue contour as before the perturbation.
677 The compensatory nature can also fit to the task demand to maintain a specific shape of
678 the tongue contour for the production of a variety of vowel sounds.

679 The observed compensatory responses worked to recover the auditory
680 characteristics (formant patterns) of the original sound. This is consistent with previous
681 experimental observations such as those of Gay et al. (1981) in which a bite-block was
682 used to perturb subjects during the articulation of English vowels resulting in
683 compensation strategies that preserved the shape of the vocal tract constriction. This
684 recovering of the original contour could be guided by a voluntary compensation
685 mechanism. Since voluntary reactions involve a cortical processing and a decision
686 process, the resultant latency for voluntary compensation is typically long. Previous
687 studies demonstrated that the reaction time of muscle activation in response to external
688 stimulations was 150 ±13 ms in the jaw (Ottenhoff et al. 1992), 315.7± 98.4ms in lips (Ito
689 et al. 2005) and 154 ms in fingers (Cole and Abbs 1987). The compensatory response
690 observed in our study was induced less than 100 ms after the perturbation onset, for the
691 first phase of the response (R2-1) and 140 ms after the perturbation for the second phase
692 (R2-2). Considering muscle contraction delays from muscle activation onset to force
693 generation (Ito et al. 2004; Mannard and Stein 1973) and the duration of the mechanical
694 response from force generation to the onset of movement, we established that the
695 latencies measured in the current compensatory responses were shorter than the latencies
696 of voluntary reaction observed in former studies. Accordingly, we conclude that the

current compensatory response does not result from a voluntary action, but from other faster mechanisms.

Given the findings in Gomi et al (2002), these faster mechanisms can be driven by passive mechanical properties and by reflexes. These two mechanisms are distinguished based on response latency. Mechanical response is the fastest response, which can be induced just after the movement change by the perturbation. We have assumed that the articulatory manner (voicing, whispering, posturing) can be controlled depending on the adjustment of mechanical properties, such as stiffness and viscosity (Burdet et al. 2001; Gomi et al. 2002; Gomi and Osu 1998). A reflex response occurs after the passive response due to neural transmission delays (Gomi et al. 2002). Depending on the pathway of neural feedback, such as via the brainstem or up to the cortex, there are several possible response latencies (Ito et al. 2005; McClean 1991). In our experimental data, we found an inflection point during compensatory response that corresponds to the acceleration peak around 140 ms after the perturbation onset. We interpreted this inflection point in displacement as the point of initiation of a reflex response, which means that the remaining part of the compensatory response (R2-2 interval in the right panel of Figure 2) was driven by reflex mechanisms. Our mathematical simulation supports this idea because the change in reflex gain in the model induced significant changes of the amplitude of the compensatory response. In the orofacial system, cortical reflexes have been shown to be involved in compensatory responses in the lips, with lip muscle responses occurring 47.5ms (± 27.5ms) after perturbation of the jaw, with the corresponding change in movement observed 100 ms after the perturbation onset (Gomi et al. 2002). Since the latency of the kinematic compensatory response in our experiment

720 was longer than the one observed in lip adjustments due to cortical reflex in jaw
721 perturbation studies, a cortical reflex could also be involved in the response of our
722 subjects.

723 Neilson et al. (1979) had failed to find clues for tonic stretch reflex of the tongue
724 with electromyographical recording from five tongue muscles (genioglossus, geniohyoid,
725 mylohyoid, styloglossus and tongue intrinsic muscle). They used a sponge forceps to grip
726 and stretch the tongue. It can be difficult to apply a force to the tongue in a consistent
727 manner with such a manual perturbation. Moreover, the force to grip the tongue with the
728 sponge forceps could have had an undesired effect and suppressed any reflexes, including
729 a stretch reflex. Contrary to this finding, we found concordant elements in the latency of
730 the experimentally observed responses, and in the similarities between simulated
731 responses with the biomechanical model and observed responses. These results support
732 the hypothesis of the existence of a reflex in the control of the tongue posture. Given the
733 nature of the current tongue perturbation, the observed compensatory response may
734 correspond directly to the stretch reflex as it would be in skeletal muscles. This can be
735 supported by the observation of the activation pattern of the styloglossus in our
736 biomechanical model. Since the neural connection and the sensory origin could not be
737 determined in the behavioral study alone, further investigation including the recording of
738 muscle activation is required to confirm the existence of stretch-like reflex in the tongue.

739 The current study can provide a direct clue about stiffness variations of the tongue
740 during speech production. By using a robotic device, we consistently applied the same
741 amount of force to the tongue in the three vowel articulations (/i/, /e/, and /ɛ/) and in the
742 three articulatory manners (voicing, whispering and posturing). Comparing the vowel

743 articulation, the smallest displacement due to the perturbation was induced in the
744 articulation of /i/ and the displacement was increased from /e/ to /ɛ/. This is consistent
745 with the previous findings described in the Introduction. The current study
746 experimentally examined the variation of tongue stiffness depending on the requirements
747 of speech utterances. Changes in tongue stiffness also explain the differences observed
748 across articulatory manners. We found that the displacement induced by the force
749 perturbation was smaller in the voicing and whispering conditions than in the posturing
750 condition. This indicates that intrinsic tongue mechanical stiffness was larger when the
751 subjects were actually producing speech sounds, and suggests that speech production
752 requires a larger amount of muscle activation. Since we do not find any difference
753 between voicing and whispering condition, the activation level of tongue muscles seems
754 not to be reliably affected by the simultaneous involvement of vocal fold vibrations.
755 Taken altogether these findings support the idea that the stiffness of the tongue can be
756 modified depending on the vowel and on the articulatory manner. Interestingly, response
757 patterns in the R2-2 interval were different from the response patterns in all the others
758 R1, R2-1 and R4 intervals; this result also supports the idea that the compensatory
759 response in R2-2 was not driven by a passive component, but rather by neural feedback
760 loop, such as a reflex.

761 Stiffness can also be different between anterior and posterior parts of the tongue.
762 We found that displacement due to the passive component of the perturbation response
763 was different across sensor locations on the tongue (TT, TB and TD). While TB and TD
764 responded with a similar amplitude of displacement, TT was smaller than the others. The
765 results indicate that the passive stiffness of the tissue connecting TB and TD is high

766 enough to synchronize their movements, while the stiffness of the connection between
767 TB and TT is relatively low. Accordingly, the passive mechanical characteristics of the
768 tongue may not be homogeneous in the tongue body. Investigating mechanical properties
769 of the tongue body is difficult because those properties may be different in tests carried
770 out on cadavers and in living subjects due to basic muscle tonus. Applying a mechanical
771 stretch to the several different sites during the same task is one way to further investigate
772 the mechanical properties of the tongue body in detail.

773 We found that the amplitude of the compensatory response in speaking tasks was
774 greater in the anterior region of the tongue than in the posterior region. This may be
775 related to the composition of the muscle fibers. (Stål et al. 2003) found that fiber
776 composition in intrinsic tongue muscles varied depending on the part of the tongue. The
777 anterior region of the tongue contained predominantly fast twitch fibers (Type II), in
778 contrast to the posterior region which contains more slow twitch fibers (Type I) than
779 Type II fibers. This difference in muscle composition may be due to the functional role of
780 the different muscles in making quick and flexible movements. Accordingly,
781 compensatory responses to external disturbances may be included in those functional
782 demands.

783 In speech motor control, auditory feedback can also be a possible loop to induce
784 compensatory movements together with somatosensory feedback (Houde and Jordan
785 1998; Lametti et al. 2012). In previous studies (Larson et al. 2000; Purcell and Munhall
786 2006), the onset of auditory compensation was found to be more than 200 ms after the
787 perturbation onset when a sustained vowel was perturbed at a certain time in pitch or
788 formant due to altered auditory feedback, which is a situation similar to the one in the

789 current study. In Purcell and Munhall (2006), when the first formant was suddenly shifted
790 by 136 Hz on average at a certain time during the sustained vowel production, the
791 auditory perturbation was gradually compensated from around 300 ms after the
792 perturbation onset, and the amplitude of compensation reached more than 20 Hz at 800
793 ms after the perturbation onset (approx. 15% of compensation). This manner of
794 compensation is different from the findings of our study, in which the acoustic changes
795 induced by the tongue perturbation were immediately compensated at the latest 300 ms
796 after the perturbation onset, without any gradual compensation afterwards. In addition,
797 the maximum change of the first formant induced by the tongue perturbation during the
798 production of /ɛ/ was approximately 55 Hz on average across subjects, which is about
799 half of Purcell and Munhall's acoustical perturbation. In contrast, the magnitude of the
800 compensation in the tongue perturbation was 27 Hz on average, which corresponds to
801 approximately 50% of the magnitude of the initial formant change. Thus, larger amounts
802 of acoustical compensation were achieved in our study than in the case of auditory
803 perturbation. Considering that the acoustical change was synchronized with the
804 compensatory movement in the tongue, we believe that the current compensation is
805 induced predominantly by somatosensory error rather than auditory error. In addition, the
806 auditory error that remained after the first compensatory response (i.e. after 250 ms) in
807 the production of /ɛ/ was not further compensated. Since the sound in perturbed condition
808 was still in the perceptual range of vowel /ɛ/, this acoustical error may be ignored to be
809 corrected, as opposed to the somatosensory error. The auditory error accompanying the
810 articulatory perturbation, as in the current study, may be easily compensated by simply
811 compensating for the articulatory changes. On the contrary the auditory error that occurs

812 with unperturbed articulation, such as in altered auditory feedback, may be difficult to
813 compensate for probably due to a conflict between auditory and somatosensory
814 requirements (Katseff et al. 2012; Lametti et al. 2012; Patri et al. 2019).

815 Although the previous studies (Larson et al. 2000; Purcell and Munhall 2006)
816 showed that the compensatory response via auditory feedback is longer than 200 ms, Cai
817 et al. (2011) also reported a relatively faster auditory compensation (approx. 120 ms).
818 Since this latency is comparable to the one in the current compensatory response,
819 auditory feedback may be involved in the production of the current compensatory
820 movement. Different from the current situation where the perturbation was applied during
821 the static articulation of the vowels, as it was in Larson et al. (2000) and Purcell and
822 Munhall (2006), Cai et al. (2011) applied auditory perturbation in a dynamical manner
823 since the first formant was altered during the production of a vowel-to-vowel gesture.
824 The arm motor control studies suggested that the latency of compensation to sensory
825 errors can be different between dynamical and static motor situations. The changes of
826 muscle activation by reflexes in response to a force perturbation occur with a longer
827 latency during a static posturing (80 ms, Soechting and Lacquaniti 1988)), than during a
828 movement (40 ms, Soechting (1988)). This suggests that the response via auditory
829 feedback in static situation can take longer than the 120 ms latency found in Cai et al.
830 (2011), and that the response to the current compensatory response occured sufficiently
831 early to reject the possility that it would be due to auditory feedback. We conclude that
832 the possible involvement of auditory feedback is unlikely in the compensation observed
833 in our study.

These converging indications concerning a major role for somatosensory feedback in the current compensatory response also suggest that correcting the somatosensory error is enough to recover the perceptually relevant characteristics of the original speech sound. Since similar speech sounds can be produced with different articulatory configurations, multiple choices for articulatory compensation can be possible for compensation of the speech sound. However, the compensatory movement observed in our study was achieved in a consistent manner across the vowels and subjects, which tends to preserve the tongue shape in the constriction region. This suggests that the speech motor control system could make use of a one-to-one auditory-somatosensory mapping in order to achieve a specific auditory goal using somatosensory feedback alone.

Grants

This work was supported by grants from the European Research Council under the European Community's Seventh Framework Program (FP7/2007-2013 Grant Agreement no. 339152), the National Institute on Deafness and Other Communication Disorders Grant R01-DC017439, and the French Ministry for Research and Education (MENRT; doctoral grant to Andrew Szabados)

References

Badin P, **Bailly G**, **Reveret L**, **Baciu M**, **Segebarth C**, **Savariaux C**. Three-dimensional linear articulatory modeling of tongue, lips and face, based on MRI and video images. *J Phon* 30: 533–553, 2002.

Beckman ME, **Jung T**, **Lee S**, **de Jong K**, **Krishnamurthy AK**, **Ahalt SC**, **Cohen KB**, **Collins MJ**. Variability in the production of quantal vowels revisited. *J Acoust Soc Am* 97: 471–490, 1995.

Brunner J, **Hoole P**, **Perrier P**. Adaptation strategies in perturbed /s/. *Clin Linguist Phon* 25: 705–24, 2011.

Buchaillard S, **Perrier P**, **Payan Y**. A biomechanical model of cardinal vowel production: muscle activations and the impact of gravity on tongue positioning. *J Acoust Soc Am* 126: 2033–51, 2009.

Burdet E, **Osu R**, **Franklin DW**, **Milner TE**, **Kawato M**. The central nervous system stabilizes unstable dynamics by learning optimal impedance. *Nature* 414: 446–449, 2001.

Cai S, **Ghosh SS**, **Guenther FH**, **Perkell JS**. Focal manipulations of formant trajectories reveal a role of auditory feedback in the online control of both within-syllable and between-syllable speech timing. *J Neurosci* 31: 16483–90, 2011.

Cole KJ, **Abbs JH**. Kinematic and electromyographic responses to perturbation of a rapid grasp. *J Neurophysiol* 57: 1498–510, 1987.

Cooper S. Muscle spindles in the intrinsic muscles of the human tongue. *J Physiol* 122: 193–202, 1953.

Feldman AG. Once more on the equilibrium-point hypothesis (lambda model) for motor control. *J Mot Behav* 18: 17–54, 1986.

Fowler CA, **Turvey MT**. Immediate compensation in bite-block speech. *Phonetica* 37: 306–26, 1981.

Gay T, **Lindblom B**, **Lubker J**. Production of bite-block vowels: acoustic equivalence by selective compensation. *J Acoust Soc Am* 69: 802–10, 1981.

Gick B, **Allen B**, **Roewer-Després F**, **Stavness I**. Speaking Tongues Are Actively Braced. *J Speech Lang Hear Res* 60: 494–506, 2017.

Gomi H, **Ito T**, **Murano EZ**, **Honda M**. Compensatory articulation during bilabial fricative production by regulating muscle stiffness. *J Phon* 30: 261–279, 2002.

Gomi H, **Osu R**. Task-dependent viscoelasticity of human multijoint arm and its spatial characteristics for interaction with environments. *J Neurosci* 18: 8965–8978, 1998.

Gottlieb GL, **Agarwal GC**. Compliance of single joints: elastic and plastic characteristics. *J Neurophysiol* 59: 937–951, 1988.

Honda M, **Kaburagi T**, **Fujino A**. Compensatory responses of articulators to unexpected perturbation of the palate shape. *J Phon* 30: 281–302, 2002.

Hoole P, **Zierdt A**. Five-dimensional articulography. *Speech Motor Control: New Developments in Basic and Applied Research* , 2010. doi:10.1093/acprof:oso/9780199235797.003.0020.

Houde JF, **Jordan MI**. Sensorimotor adaptation in speech production. *Science* 279: 1213–6., 1998.

Ito T, **Gomi H**. Cutaneous mechanoreceptors contribute to the generation of a cortical reflex in speech. *Neuroreport* 18: 907–10, 2007.

Ito T, **Kimura T**, **Gomi H**. The motor cortex is involved in reflexive compensatory adjustment of speech articulation. *Neuroreport* 16: 1791–4, 2005.

Ito T, **Murano EZ**, **Gomi H**. Fast force-generation dynamics of human articulatory muscles. *J Appl Physiol* 96: 2318–24; discussion 2317, 2004.

Katseff S, **Houde J**, **Johnson K**. Partial compensation for altered auditory feedback: a tradeoff with somatosensory feedback? *Language and speech* 55: 295–308, 2012.

Keating PA. The window model of coarticulation: articulatory evidence. In: *Papers in Laboratory Phonology*. Cambridge University Press, 1990, p. 451–470.

Kelso JAS, **Tuller B**, **Vatikiotis-Bateson E**, **Fowler CA**. Functionally specific articulatory cooperation following jaw perturbations during speech: evidence for coordinative structures. *J Exp Psychol Hum Percept Perform* 10: 812–832, 1984.

Kier WM, **Smith KK**. Tongues, tentacles, and trunks: The biomechanics of movement in mucular-hydrostats. *Zool J Linn Soc* 83: 307–324, 1985.

Lametti DR, **Nasir SM**, **Ostry DJ**. Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *J Neurosci* 32: 9351–8, 2012.

Larson CR, **Burnett TA**, **Kiran S**, **Hain TC**. Effects of pitch-shift velocity on voice F0 responses. *J Acoust Soc Am* 107: 559–64, 2000.

MacNeilage PF. Motor control of serial ordering of speech. *Psychol Rev* 77: 182–196, 1970.

Mannard A, **Stein RB**. Determination of the frequency response of isometric soleus muscle in the cat using random nerve stimulation. *J Physiol* 229: 275–296, 1973.

Marsden CD, **Merton PA**, **Morton HB**. Servo action in human voluntary movement. *Nature* 238: 140–3, 1972.

McClean MD. Lip muscle reflex and intentional response levels in a simple speech task. *Exp Brain Res* 87: 662–70, 1991.

McFarland DH, **Baum SR**, **Chabot C**. Speech compensation to structural modifications of the oral cavity. *J Acoust Soc Am* 100: 1093–104, 1996.

Nazari MA, **Perrier P**, **Chabanas M**, **Payan Y**. Shaping by stiffening: a modeling study for lips. *Motor Control* 15: 141–168, 2011.

Nazari MA, **Perrier P**, **Payan Y**. The distributed lambda (λ) model (DLM): a 3-D, finite-element muscle model based on Feldman's λ model; assessment of orofacial gestures. *J Speech Lang Hear Res* 56: S1909-1923, 2013.

Neilson PD, **Andrews G**, **Guitar BE**, **Quinn PT**. Tonic stretch reflexes in lip, tongue and jaw muscles. *Brain Res* 178: 311–27, 1979.

Ottenhoff FA, **van der Bilt A**, **van der Glas HW**, **Bosman F**. Peripherally induced and anticipating elevator muscle activity during simulated chewing in humans. *J Neurophysiol* 67: 75–83, 1992.

Patri J-F, **Diard J**, **Perrier P**. Optimal speech motor control and token-to-token variability: a Bayesian modeling approach. *Biol Cybern* 109: 611–626, 2015.

Patri J-F, **Diard J**, **Perrier P**. Modeling Sensory Preference in Speech Motor Planning: A Bayesian Modeling Framework. *Front Psychol* 10, 2019.

Payan Y, **Perrier P**. Synthesis of V-V sequences with a 2D biomechanical tongue model controlled by the Equilibrium Point Hypothesis. *Speech Commun* 22: 185–205, 1997.

Perkell JS, **Cohen MH**, **Svirsky MA**, **Matthies ML**, **Garabieta I**, **Jackson MT**. Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. *J Acoust Soc Am* 92: 3078–3096, 1992.

Perkell JS, **Nelson WL**. Variability in production of the vowels /i/ and /a/. *J Acoust Soc Am* 77: 1889–1895, 1985.

Perrier P, **Payan Y**, **Zandipour M**, **Perkell J**. Influences of tongue biomechanics on speech movements during the production of velar stop consonants: a modeling study. *J Acoust Soc Am* 114: 1582–99, 2003.

Pruszynski JA, **Scott SH**. Optimal feedback control and the long-latency stretch response. *Exp Brain Res* 218: 341–359, 2012.

Purcell DW, **Munhall KG**. Compensation following real-time manipulation of formants in isolated vowels. *J Acoust Soc Am* 119: 2288–97, 2006.

Rabiner LR, **Schafer RW**. *Digital processing of speech signals*. Prentice-Hall, 1978.

Savariaux C, **Perrier P**, **Orliaguet JP**. Compensation Strategies for the Perturbation of the Rounded Vowel [U] Using a Lip Tube - a Study of the Control Space in Speech Production. *J Acoust Soc Am* 98: 2428–2442, 1995.

Schmidt RA, **Lee TD**. *Motor Control And Learning: A Behavioral Emphasis*. Human Kinetics, 2005.

Soechting JF. Effect of load perturbations on EMG activity and trajectories of pointing movements. *Brain Res* 451: 390–396, 1988.

Soechting JF, **Lacquaniti F**. Quantitative evaluation of the electromyographic responses to multidirectional load perturbations of the human arm. *J Neurophysiol* 59: 1296–1313, 1988.

Stål P, **Marklund S**, **Thornell LE**, **De Paul R**, **Eriksson PO**. Fibre composition of human intrinsic tongue muscles. *Cells Tissues Organs* 173: 147–61, 2003.

Stone M, **Lundberg A**. Three-dimensional tongue surface shapes of English consonants and vowels. *J Acoust Soc Am* 99: 3728–3737, 1996.

Takemoto H. Morphological analyses of the human tongue musculature for three-dimensional modeling. *J Speech Lang Hear Res* 44: 95–107, 2001.

Westbury JR. *X-Ray Microbeam Speech Production Database User's Handbook*. Waisman Center on Mental Retardation & Human Development, 1994.

Whalen DH, **Iskarous K**, **Tiede MK**, **Ostry DJ**, **Lehnert-Lehouillier H**, **Vatikiotis-Bateson E**, **Hailey DS**. The Haskins optically corrected ultrasound system (HOCUS). *J Speech Lang Hear Res* 48: 543–53, 2005.

Xing F, **Woo J**, **Murano EZ**, **Lee J**, **Stone M**, **Prince JL**. 3D tongue motion from tagged and cine MR images. *Med Image Comput Comput Assist Interv* 16: 41–48, 2013.

Figure Legends

Figure 1: A: Experimental setup of tongue perturbation. Green points represent the location of sensors of electromagnetic articulograph [TT: tongue tip, TB: tongue blade, TD: tongue dorsum, UL: upper lip, LL: lower lip, J: jaw, and reference makers (nasion, right and left mastoid and upper incisor)]. B: the view of tongue surface representing the locations of EMA sensors (TT, TB and TD) and anchor points for the tongue perturbation (TP). C: Schematic view of the 2D tongue model. The nodes on the tongue surface are numbered from 1 (tongue root) to 16 (tongue tip). The nodes, 11, 13 and 16 were chosen as the corresponding location of measured sensors (TD, TB, and TT, respectively). The four nodes (nodes 13 and 14 and two nodes below those) represented by blue-filled points were chosen as the area for the tongue perturbation (TP). Muscle representations of styloglossus and posterior genioglossus were shown as representative examples. Styloglossus is shaded in red and posterior genioglossus is in light blue. The bold thick line (styloglossus in black and posterior genioglossus in gray) represents macrofibers and the shaded area represents the elements whose mechanical stiffness was changed by muscle activation.

Figure 2: A: Horizontal displacement change in the jaw (J), upper and lower lips (UL and LL), and tongue tip, blade and dorsum (TT, TB and TD). The shaded area represents the standard error across the subject. The vertical dashed and dotted lines represent remarkable inflection points in the trajectory that was detected in TB. B: The magnified view of the horizontal displacement, horizontal velocity and horizontal acceleration in

TB. The time range corresponds to the range shown with the thick horizontal bar in A. The vertical lines also correspond to the ones in A.

Figure 3: A: Trajectory of tongue blade (TB) displacement due to the perturbation from the onset of the perturbation (circle mark) to the offset (x mark) in voicing condition. Each line represents three vowel conditions. * marks represent a detected local peak of the movement. + marks represent acceleration peaks during compensatory response. The estimated tongue surfaces in each vowel condition are superimposed as gray lines. B: Averaged displacements in each phase. Top panel: grouping according to the vowels; Middle panel: grouping according to articulatory manners; Bottom panel: grouping according to the sensor locations. Error bars represent standard error.

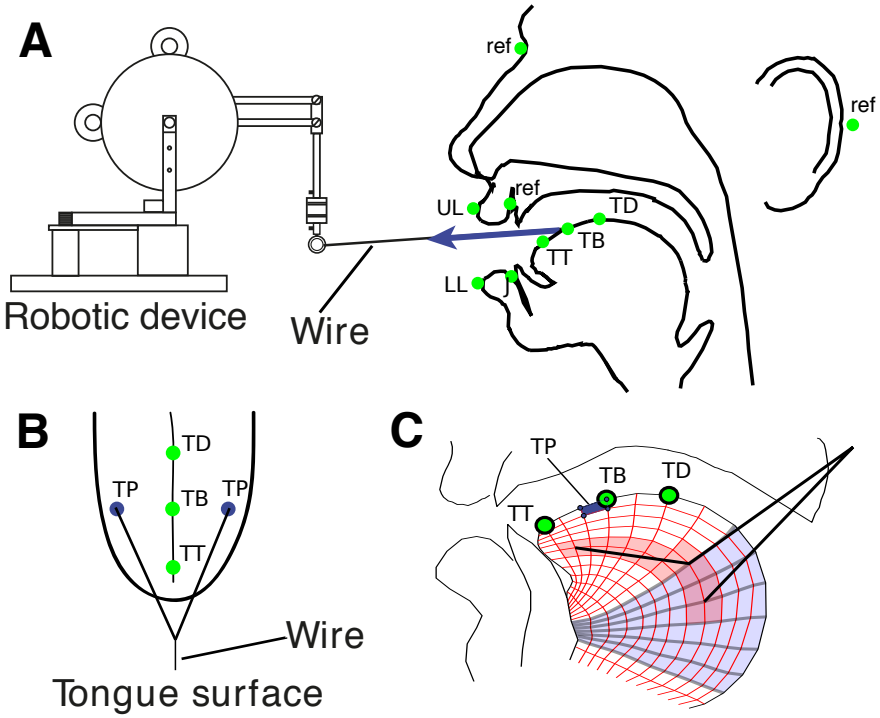
Figure 4: Displacements of compensatory response (R2-2 in Figure 2). Error bars represent standard error.

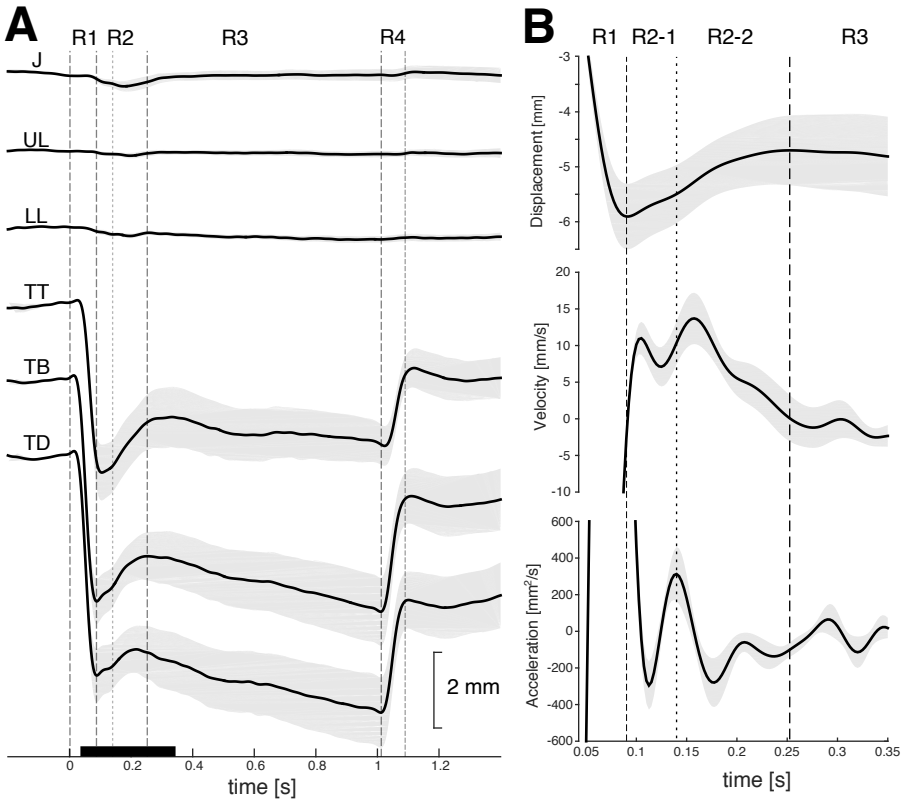
Figure 5: Temporal patterns of the first, second and third formants in speech sounds. Triangles represent the temporal points corresponding to the inflection points in displacement, which is represented as the vertical line in Figure 2. The shaded area represents the standard error across participants.

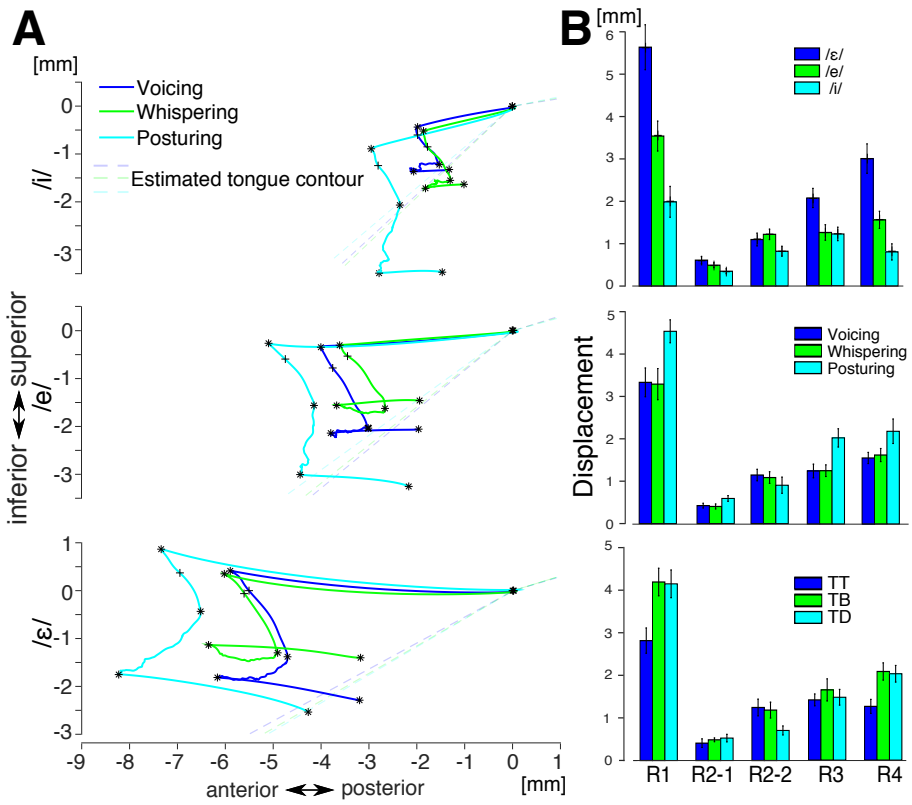
Figure 6: Simulated results. A: time response in horizontal displacement of TB with each reflex gain. The vertical lines represent the peak points of the first initial response (R1) and of the compensatory responses (R2). B: TB displacement in the sagittal plane. C: The

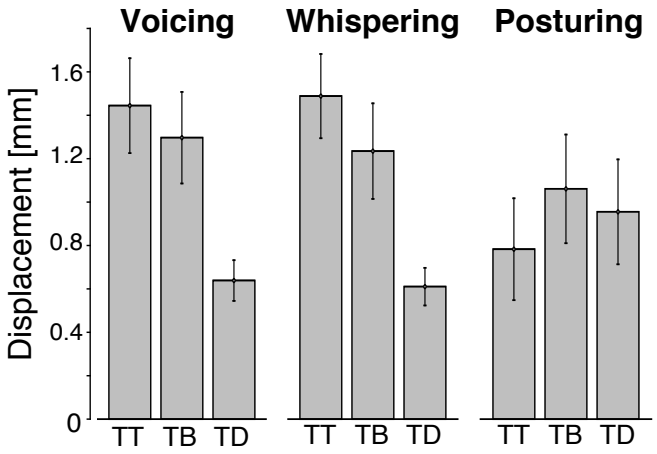
peak displacement of R1 in three tongue sites (TT, TB and TT). D: The peak amplitude of compensatory response (R2) in three tongue sites.

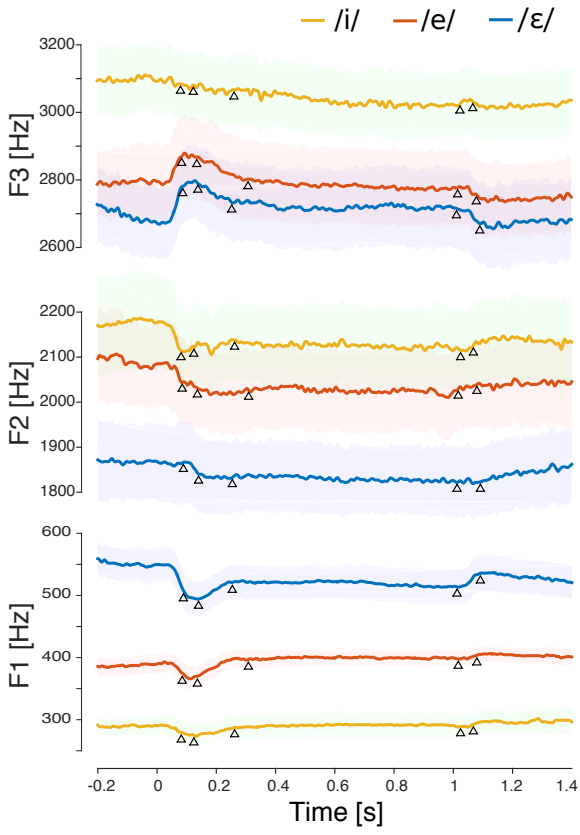
Figure 7: Simulated results. A: time response of muscle activation level. The amplitude of muscle activation was represented in local unit of the simulation. B: base line muscle activity for the production of /e/ in simulation.

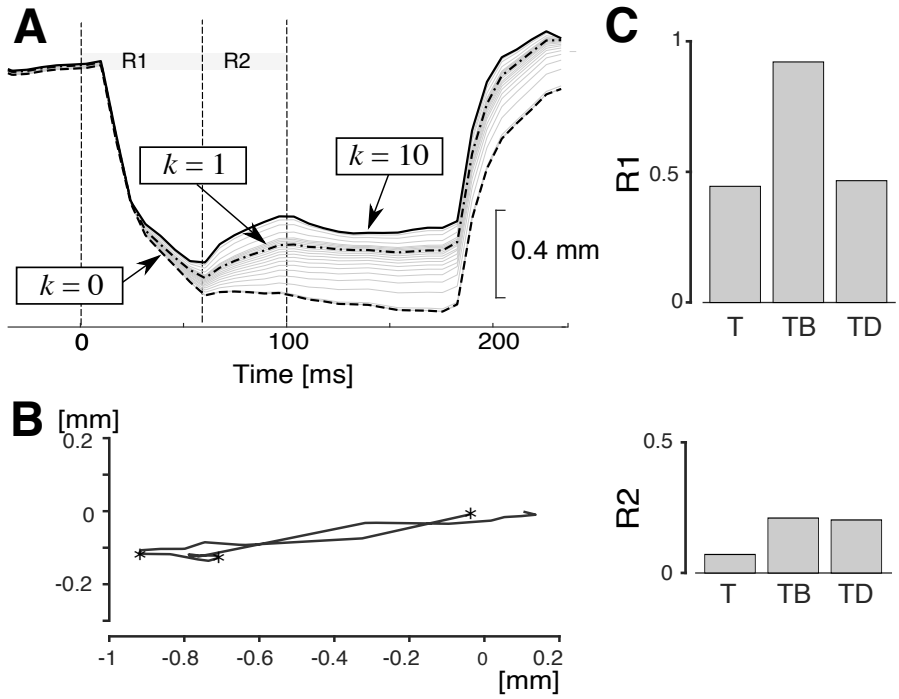

A
ref
ref
ref
UL
TD
TB
TT
LL
Robotic device
Wire
B
TD
TP
TB
TP
TT
Wire
Tongue surface
C
TP
TB
TD
TT

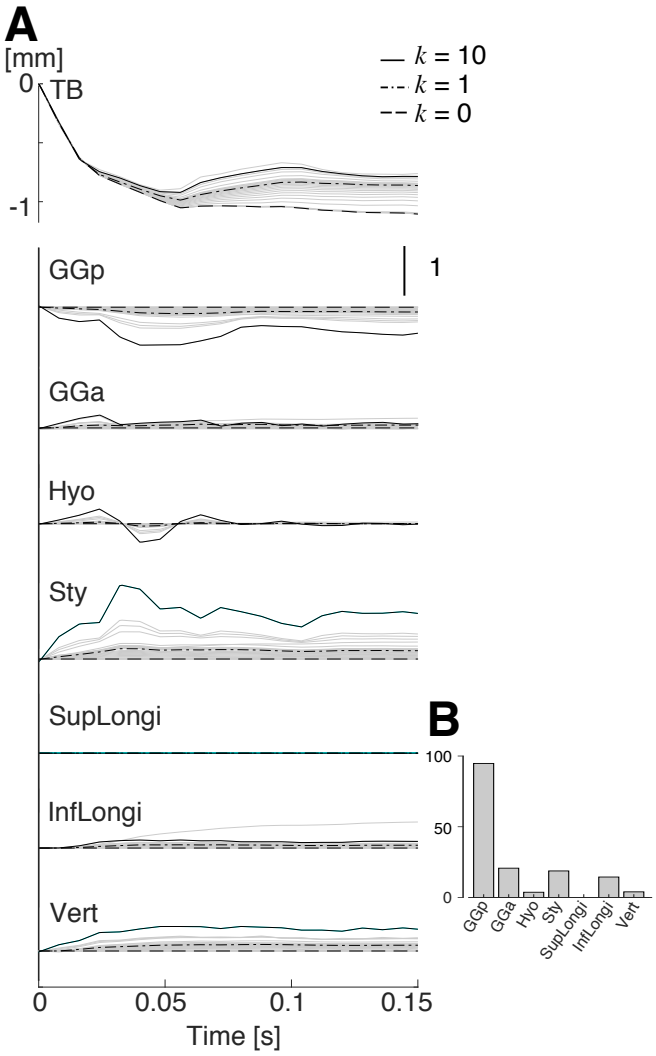

A
R1
R2
R3
R4
J
UL
LL
TT
TB
TD
2 mm
0
0.2
0.4
0.6
0.8
1
1.2
time [s]
B
R1
R2-1
R2-2
R3
Displacement [mm]
-3
-4
-5
-6
Velocity [mm/s]
20
15
10
5
0
-5
-10
Acceleration [mm²/s]
600
400
200
0
-200
-400
-600
0.05
0.1
0.15
0.2
0.25
0.3
0.35
time [s]


A
[mm]
Voicing
Whispering
Posturing
Estimated tongue contour
/i/
/e/
/ɛ/
inferior
superior
anterior
posterior
[mm]
B
[mm]
/ɛ/
/e/
/i/
Voicing
Whispering
Posturing
Displacement
TT
TB
TD
R1
R2-1
R2-2
R3
R4


Voicing
Whispering
Posturing
Displacement [mm]
1.6
1.2
0.8
0.4
0
TT
TB
TD
TT
TB
TD
TT
TB
TD


/i/
/e/
/ɛ/
F3 [Hz]
3200
3000
2800
2600
F2 [Hz]
2200
2100
2000
1900
1800
F1 [Hz]
600
500
400
300
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
1.4
Time [s]


A
R1
R2
k = 1
k = 10
k = 0
0.4 mm
0
100
200
Time [ms]
B
[mm]
0.2
0
-0.2
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
[mm]
C
R1
1
0.5
0
T
TB
TD
R2
0.5
0
T
TB
TD


A
[mm]
0
TB
-1
k = 10
k = 1
k = 0
GGp
1
GGa
Hyo
Sty
SupLongi
InfLongi
Vert
0
0.05
0.1
0.15
Time [s]
B
100
50
0
GGp
GGa
Hyo
Sty
SupLongi
InfLongi
Vert