Posture stabilization of the tongue for speech: responses to mechanical perturbation

Takayuki Ito, Jean-Loup Caillet, Pascal Perrier

Posture stabilization of the tongue for speech: responses to mechanical perturbation

Takayuki Ito[1,2], Jean-Loup Caillet [3], Pascal Perrier[1]

1: GIPSA lab, CNRS, Grenoble-INP, Univ. Grenoble Alps, 2: Haskins Laboratories,
3: CS System of information
takayuki.ito@gipsa-lab.grenoble-inp.fr

ABSTRACT

The control of tongue movements is crucial in speech production, since the tongue is responsible for the fine shaping of the vocal tract, which determines spectral characteristics of the speech sounds. Little is however known about motor control mechanisms of the tongue because of the technical difficulty to have physical access to the tongue. The study aims to investigate a control mechanism of the tongue for posture stabilization for speech. We newly developed mechanical perturbation system to change the tongue posture during speaking and examined responses to transient mechanical perturbations during speaking. We found that vocal tract shape was stabilized by compensatory responses of the tongue for vowel production. The response was faster than auditory feedback loops and tended to maintain the shape of tongue contour. The amplitude of response is also greater for voiced and whispered vowel than for posturing.

Keywords: Vowel production, Sensorimotor control, Compensatory mechanism, Reflex, Somatosensation.

1. INTRODUCTION

Tongue is known to be the most important articulator in speech production. Their control has to be fine and accurate to determine spectral characteristics in speech sound that are relevant for an efficient phonetic categorization in listeners. This great accuracy is achieved under many different physical conditions, such as standing, lying or running. Yet, tongue is a complex muscular hydrostat [5] that deforms non-linearly as a consequence of the activation of more than 20 muscles [1,8]. In this context, a key question for better understanding of speech production is what neurophysiological mechanisms are involved in the achievement of accurate and stable motor control and stabilization. The previous studies showed that saturation effects [3] combined with palatal contacts [1] could help stabilizing tongue postures in high vowels and in palatal consonants. More generally Gick et al. [4] have proposed that mechanical properties (stiffness and viscosity) of the tongue would be systematically controlled in order to make constrictions properly with the appropriate accuracy. As an alternative explanation, which is compatible with the suggestions summarized above but would apply more generally to the control of tongue movement in all its biological functions (speaking, swallowing, breathing), the control due to somatosensory feedback would enable achieving and maintaining postures in a number of external physical conditions, just as stretch reflex in the control of arm or leg postures [6]. Stretch reflex is known to be mediated by muscle spindles. While there is an evidence for the existence of muscle spindles in the tongue [2], the muscle spindles were not found in each tongue muscle and it has been suggested that their density could be low. Hence, it is not clear whether they could significantly contribute to the postural control of the tongue. However, other short delay somatosensory reflexes could also originate in the tongue from other mechanoreceptors [9], which are empirically known by feeling and guiding of the food bolus during swallowing.
The current study was designed to find behavioral evidence concerning a short delay reflex in the tongue and its role for the stabilization of the tongue during speech production, without investigating the exact nature of the sensory receptors. More specifically this study investigated on-line correction mechanisms of tongue control by an external perturbation during the production of speech sounds. We developed an original mechanical perturbation system to the tongue that externally change the tongue posture. We examined (1) whether or not the tongue showed quick compensatory responses to correct movement errors due to the external perturbation and (2) how the response induced by the perturbation varied according to whether the motor task is a speech or a non-speech one.

2. METHOD

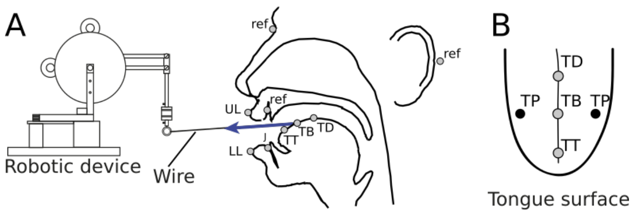
2.1. Participants

Eight native French speakers participated in the experiment. The participants were all healthy young adults with normal hearing. All procedures were approved by the ethical committee in University of Grenoble Alps.[1]

2.2. Data recording

The displacements of the tongue, lip and jaw were recorded using a 3D electromagnetic articulograph (Wave, Northern Digital Inc.). The sensors were glued in the mid-sagittal plane, on the tongue tip (TT), tongue blade (TB), and tongue dorsum (TD), on the upper and lower lips (UL and LL), and on the jaw (J) (see Fig. 1). Four reference points, namely the nasion, the left and right mastoids and the upper incisor were also recorded for head movement correction. The data were recorded at 400 Hz sampling rate. The speech signal was synchronized with the movement data and was recorded at a 22.05 kHz sampling frequency.

Figure 1: A, Tongue perturbation system (left) and EMA sensor locations in the mid-sagittal plane; B: EMA sensor locations and anchor points (TP) of the wire on the tongue surface.



2.3. Tongue Perturbation

A small robotic device (Phantom, 3D Systems) was used to apply a load to the tongue (see Fig. 1A). The robot was set in front of the subject and connected to the tongue surface through thin wire. The wire has two small anchors that were glued on the tongue surface lateral to the middle sensor (TPs in Fig. 1B). The tongue perturbation was produced by pulling the tongue with a force step of 1N that held for 1 second. The interference of this additional anchors to speech production is relatively small compared with the sensors of the articulography because the height of this anchor is relatively low (approx. 1mm).

2.4. Experimental Procedure

The participant sat on the chair with head holder. The robot and display were set in front of the participant. The task consisted of the production of vowels /i/, /e/, and /ɛ/ under 3 conditions: (1) in voiced speech (Voicing); (2) in whispered speech (Whispering); (3) silently, i.e. maintaining tongue postures of the task vowels (Posturing). These tasks were carried out in random order. Written visual instructions were used to inform the participants about the task for the next trial. The experimenter controlled the start of each trial by examining if the participant was ready to speak. The visual onset cue for speaking were presented 0.5 s after the trial onset. The tongue perturbation was applied 1.5 s after the trial onset in only one third of the trials selected randomly in order to avoid any anticipation of the presence/absence of the perturbation.

2.5. Data Analysis

For the articulatory movement data, a 3D head movement correction was applied off-line based on the head position measured using the 4 reference sensors. Jaw displacement was subtracted from the displacements in the tongue sensors in order to focus on tongue control only. We computed the average of displacement amplitude induced by the perturbation, which was calculated taking the tongue position at perturbation onset as “zero-displacement” reference. For rough estimation of the average tongue contour of the production of the task vowels, direction vectors joining between each tongue sensors, from TB to TT and from TB to TD, were calculated in each participant and was averaged across participants.

The acoustical analysis was carried out on the voiced speech only. The first and second formants (F1 and F2) were extracted using LPC analysis [7]. These formant values have been more specifically analysed at a number of selected time points that we have considered to be relevant from the perspective of the correction mechanism.

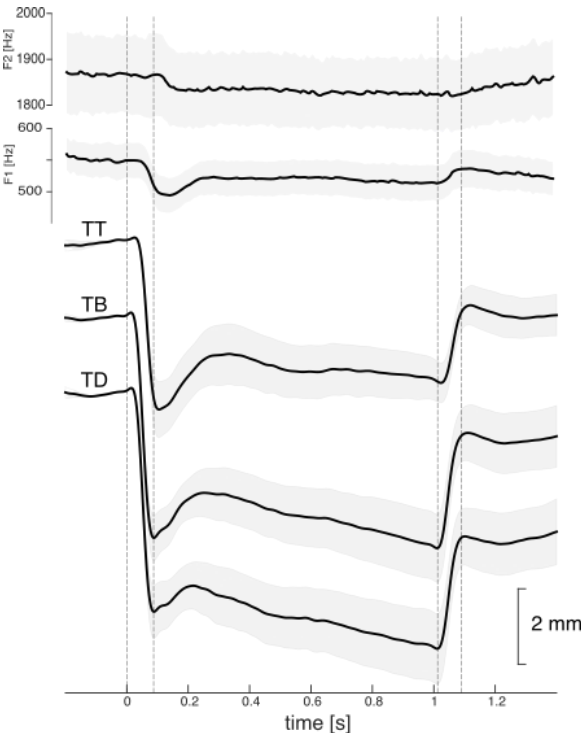
In overall statistical analyses, mixed-effect model was computed with vowels (/i/, /e/ and /ɛ/), speaking manner (Voicing, Whispering, and Posturing) and sensor locations on the tongue (TT, TB, and TD) as fixed effects, and subjects as random effects. Post-hoc tests with Bonferroni correction followed.

3. RESULTS

The mechanical perturbation disturbed the tongue posture in the horizontal direction during the speech tasks. Figure 2 shows the time variations of the average horizontal tongue sensors’ displacements and of the F1 and F2 values for vowel /ɛ/ in voiced condition. The downward displacements correspond to a forward movement of the tongue. Dashed vertical lines indicates time events on the TB variation.

The perturbation Onset (between 1st and 2nd lines) induced a large tongue displacement in the forward direction. This displacement was partially reduced by compensatory movement. After reaching a maximum of compensatory movement, the tongue was again gradually drifted forward. When the perturbation was removed (perturbation Offset), the tongue moved backwards, but it not completely recovered its original position along the horizontal direction (between the 3rd and 4th line). This sequence of movements indicates that the tongue posture was controlled efficiently to face the transient force change, but not the constant level of force maintained afterwards. This could be due to the fact that transient changes are experienced in daily life, in situation such as biting or eating, while static forces of this intensity are rather rare.

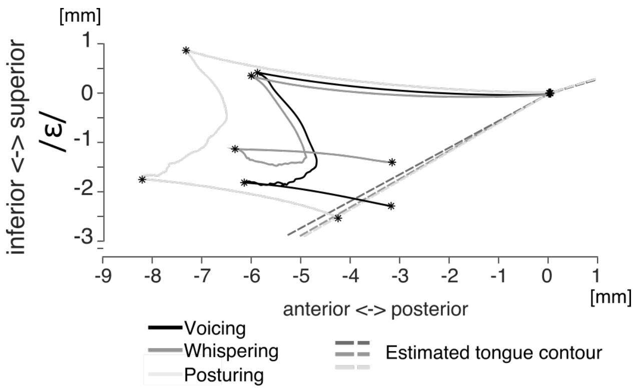
Figure 2: Temporal patterns of the first and second formant frequencies and horizontal displacements of the tongue tip (TT), blade (TB) and dorsum (TD) during the production of voiced /ɛ/. Time 0 corresponds to the perturbation onset.



The produced sound was also modified and the spectral changes were largely corrected synchronously with tongue movement corrections. The acoustical change was mostly found in F1. The peak amplitude was reliably different from the base amplitude before the perturbation ($p < 0.005$), but this was not found in F2 ($p > 0.3$). The compensation on F1 starts around 160 ms after the perturbation onset. In average approximately 50 % of the change in F1 was corrected, but F1 is still different from its value at perturbation onset ($p < 0.01$). Similar responses were also found for the production of /i/ and /e/. These results indicate that tongue position is controlled to quickly correct for the acoustical changes induced by the perturbation and maintain the integrity of the produced vowel sound.

To see spatial nature of the responses to the perturbation, the averaged trajectories of the TB sensor during the production of /ɛ/ are shown in the sagittal plane for the three conditions (Fig. 3). We spatially aligned the data at the onset of the tongue perturbation to zero. The time events indicated by the dashed lines in Fig. 2 are represented by asterisks in Fig. 3. In general, the sensor trajectories from the onset to the offset of the perturbation are very similar across conditions.

Figure 3: Displacement of tongue in sagittal plane during the production of /ɛ/ in the three conditions.

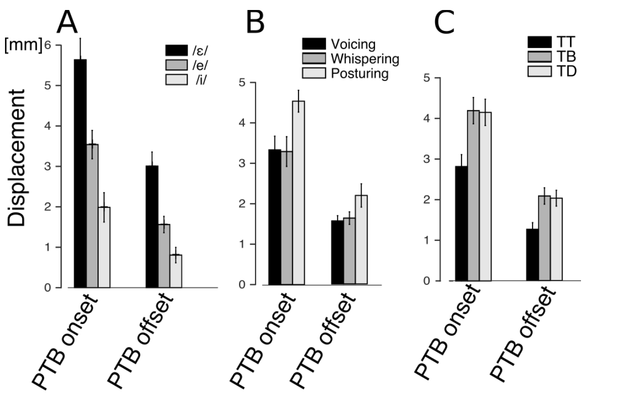


The remarkable finding is that the immediate compensatory movement back toward the original position did not follow the same path as the movement induced by the perturbation. Instead, the compensatory response seems to enable a recovering of the original tongue contour (light lines in Fig. 3). The direction of this compensatory response was nearly perpendicular to the original tongue contour. This tendency was seen in all conditions. This suggests that the immediate compensatory mechanism of the tongue is not intended to maintain the position of the tongue at a specific location, but to maintain the same shape of the tongue contour. This statement is also supported by the observation of the tongue positions after the perturbation removal, which were not the exact original position, but the positions that preserved the original shape of the tongue contour.

The Mixed-Effect Linear Model revealed a significant influence of each fixed effect, i.e. vowel, speaking manner and sensor location, separately. The movement amplitudes of the transient response following perturbation onset and at perturbation offset are presented in Fig. 4. The transient displacement amplitude following the perturbation onset is reliably different according to the vowel (Fig. 4A), [$F(2,182) = 175.5$, $p < 0.001$]. The largest displacement was found for /ɛ/ and it was gradually reduced for /e/ and /i/ production. Post-hoc tests showed significant differences in all combination of comparison ($p < 0.02$). The same tendency was found at the perturbation offset [$F(2,182) = 142.2$, p<

0.001]. This is probably due to differences in tongue stiffness depending on the vowel. The highest stiffness produced the smallest position change.

Figure 4: Displacement of tongue following the onset of the perturbation between 1st and 2nd dashed lines in Fig. 2 and at perturbation offset between 3rd and 4th dashed lines.



Differences in tongue stiffness may also explain differences in tongue response amplitude associated with the speaking manner (Fig. 4B). We found that displacements in the posturing task were reliably greater than those in the other two tasks (whispering and voicing) [$F(2,182) = 24.01, p< 0.001$], suggesting a lower stiffness in the posturing task. A similar tendency was observed at the perturbation offset, although it did not reach at significance level [$F(2,182) = 1.774, p > 0.15$]. These results support the idea that the tongue stiffness can vary depending on the speaking manner.

Finally, we found a significant difference according to the sensor location on the tongue [$F(2,182) = 26.22$, $p < 0.001$] in Fig. 4C. The transient displacement amplitude for the tongue blade (TB) was not different from the one for the tongue dorsum (TD) ($p > 0.9$), but the amplitude for tongue tip (TT) was significantly smaller than for the other two sensors ($p < 0.001$ in both). This was also observed at the perturbation offset [$F(2,182) = 22.74$, $p < 0.001$]. The reduced movement amplitude for TT may suggest that the stiffness of the anterior part of the tongue may be lower than in posterior part, which suggest that the passive mechanical characteristics is not homogeneous in the tongue body.

4. DISCUSSION

We observed an immediate compensatory response against the mechanical perturbation. The position of each sensor on the tongue did not return to their original position, but the response tends to enable recovering the original shape of the tongue contour. During the vowel production, the spectral characteristics of the sounds were also modified, but were in large part recovered quickly synchronously with the immediate compensatory response in motion, which suggests that tongue control was organized so as to maintain the acoustic output. The amplitude of the displacement due to the tongue perturbation varied depending on the speaking manners (smaller for voicing and whispering than for silent speech) presumably due to a change of tongue stiffness depending on the task demand. The current finding suggests that the tongue is controlled precisely by rapid compensatory mechanisms and impedance control to stabilize vocal tract shape for speaking.

Auditory feedback can be a possible loop to induce compensatory movements together with somatosensory feedback. When the produced sound is suddenly changed in pitch or formant by using altered auditory feedback system, those acoustical change can be compensated, in some situations, more than 200 ms of latency. In the current test, the perturbed acoustical change started around 50 ms and the compensation was seen around 160 ms after the perturbation onset, indicating the latency from the detection of acoustical change to the compensation was about 110 ms that is relatively faster than the previous compensation due to auditory error. Considering that the acoustical change was synchronized with articulatory movement change in the tongue and the change of articulatory movement was seen less then 100 ms after the perturbation onset, control by somatosensory input rather than auditory inputs may be more dominant in the current compensatory response. The result indicates that, while the tongue posture was stabilized to maintain certain acoustical goal for speaking, actual compensatory response can be driven not on an auditory-basis, but rather on a somatosensory-basis. This suggests that fast somatosensory feedback in speech production is adjusted in order to preserve acoustics, probably as a result of a learning of auditory-somatosensory mapping.

5. ACKNOWLEDGEMENTS

This work was partly supported by the European Research Council under the European Community's Seventh Framework Program (FP7/2007-2013 Grant Agreement no. 339152).

6. REFERENCES

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

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

[3] Fujimura, O., Kakita, Y. 1979. Remarks on quantitative description of lingual articulation, In: Lindblom, B., Öhman, S., (eds.), *Frontiers of Speech Communication Research*, London: Academic Press, 17–24.

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

[5] Kier, W. M., Smith, K. K. 1985. Tongues, tentacles, and trunks: The biomechanics of movement in mucular-hydrostats. *Zool J Linn Soc* 83, 307–24.

[6] Mussa-Ivaldi, F. A., N. Hogan, E. Bizzi. 1985. Neural, mechanical, and geometric factors subserving arm posture in humans. *J Neuroscience* 5, 2732–43.

[7] Rabiner, L. R., Schafer, R. W. 1978. Digital processing of speech signals. Prentice-Hall.

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

[9] Trulsson, M., Essick, G. K. 1997. Low-threshold mechanoreceptive afferents in the human lingual nerve. *J Neurophysiol* 77, 737–48.

---

[1] CERNI: Comité d'Ethique pour les Recherches Non Interventionnelles.