



HAL
open science

Compensatory response to tongue perturbation occurs similarly with normal and altered auditory feedback

Morgane Bourhis, Yosra Jelassi, Christophe Savariaux, Pascal Perrier,
Takayuki Ito

► To cite this version:

Morgane Bourhis, Yosra Jelassi, Christophe Savariaux, Pascal Perrier, Takayuki Ito. Compensatory response to tongue perturbation occurs similarly with normal and altered auditory feedback. ISSP 2024 - 13th International Seminar on Speech Production, May 2024, Autrans, France. hal-04717030

HAL Id: hal-04717030

<https://hal.science/hal-04717030v1>

Submitted on 25 Oct 2024

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

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

Compensatory response to tongue perturbation occurs similarly with normal and altered auditory feedback

Bourhis M., Jelassi Y., Savariaux C., Perrier P., Ito T.

Univ. Grenoble Alpes, CNRS, Grenoble-INP, GIPSA-lab

Morgane.bourhis@grenoble-inp.fr, takayuki.ito@gipsa-lab.grenoble-inp.fr

Abstract

Somatosensory and auditory feedback contribute to speech motor control, but it is unclear how they interact in on-line feedback control. In previous studies, we showed evidence for a somatosensory-based response to tongue-stretch perturbation in vowel production, which ensures tongue posture stabilization and preserves the auditory characteristics of the sound. In this study, we combined the tongue perturbation with an alteration of the auditory feedback, which induced formant shifts that were either consistent or inconsistent with the auditory impact of the tongue perturbation. We investigated how the compensation for the auditory perturbation interacts with the somatosensory response to the tongue perturbation. We did not find any interaction. The latency of the compensation for the formant shift was longer than the one of the somatosensory responses, suggesting that somatosensory feedback control could be the fastest one to preserve crucial auditory characteristics of vowels.

Keywords: *speech motor control, on-line feedback mechanism, mechanical perturbation, reflex*

Introduction

Speech is auditory in nature. Hence, auditory feedback is crucial to achieve speech goals and precise speech production (Savariaux et al., 1999; Perkell et al., 2000; Purcell & Munhall, 2006; Cai et al., 2011). However, somatosensory feedback has also been shown to play an important role both for speech motor control, (Tremblay et al, 2003, Nasir & Ostry, 2008) and for vowel identification in the absence of auditory feedback (Patri et al., 2020).

In a recent study (Ito et al 2020), using a sudden tongue-stretch perturbation during steady-state vowel production, we have found clear evidence for a quick on-line compensatory response (with a 130-ms latency) aiming at preserving the production of the vowel against the perturbation. We have also shown (Bourhis et al., submitted) that this compensatory response occurs similarly when the participants receive their normal auditory feedback and when their auditory feedback is masked by a pink noise. This result suggests a crucial role of somatosensory feedback in the generation of the observed response to the tongue-stretch perturbation. This was confirmed under the same experimental conditions by an EMG study of the muscles acting on the anterior part of the tongue: an increase of muscle activity was observed around 60ms after perturbation onset. This is a relatively short latency which is more compatible with polysynaptic somatosensory reflex, than with typical phonetic auditory correction (Ito et al., 2024).

Importantly, we observed that the compensatory response did not bring the tongue back to its position before the perturbation onset, but to another position that preserved the tongue contour in the constriction of the vocal tract and was compatible with the achievement of the crucial auditory characteristics of the vowel. This suggests that in speech production somatosensory feedback could be specifically tuned, so as to ensure accurate acoustic vowel production, even in the absence of auditory monitoring.

However, our results do not discard a possible role of auditory feedback, when it is available. Indeed, the condition of our experiment may not allow to demonstrate this contribution, since the somatosensory correction and the auditory correction act in the same direction, aiming at recovering the auditory characteristics of the produced vowel. Previous studies from the literature, using on-line alterations of the auditory feedback, both at the levels of the formants (Purcell & Munhall, 2006) and of the pitch (Larson et al, 2000) during steady-state vowel production, have shown latencies of the auditory correction that were longer (>200ms) than the latency of the response observed in our study (130ms). However, latencies of auditory corrections as short as 120ms were found when the perturbation was applied during the production of a sequence of vowels, i.e. under dynamical speech production conditions (Cai et al 2011, Xu et al, 2004, Donath et al 2002). Since our tongue-stretch perturbation induces a displacement of the tongue during vowel production, we cannot discard the possibility that an auditory correction mechanism associated with dynamical speech production, could also be involved in our steady-state production task.

In the condition of our study using tongue-stretch perturbation, in order for us to be able to detect the specific auditory contribution to the response of the tongue-stretch perturbation, it is necessary to break the compatibility between the somatosensory-based and the auditory-based corrections. This can be done by examining whether an auditory correction induced by altered auditory feedback is not similar to the somatosensory-based correction.

To address this, we carried out a somatosensory-auditory perturbation test by combining the tongue-stretch perturbation with an auditory perturbation. The auditory perturbation was applied to the first formant, and was either in the same direction as the acoustic consequence of the tongue-stretch perturbation (decrease of F1) or in the opposite direction (increase of F1). Based on the latencies found in the literature in steady-state vowel production, we predicted that the latency of the auditory-based correction could be longer than the one of the somatosensory corrections. Hence, the additional auditory error induced by the F1-shifts could not affect the quick compensatory response to the tongue-stretch perturbation. To verify the latency of auditory correction, we also tested auditory-perturbation alone conditions, which was applied to the first formant in both directions.

Method

Twelve native French speakers participated in the experiment. They reported no known speech or hearing impairment and no history of profound injury that could induce a somatosensory loss in the orofacial region. This experiment was approved by the local ethical committee (CERGA: Comité d'éthique pour la recherche, Grenoble-Alpes [CERGA-AvisConsultatif-2021-18]). All participants signed the consent form.

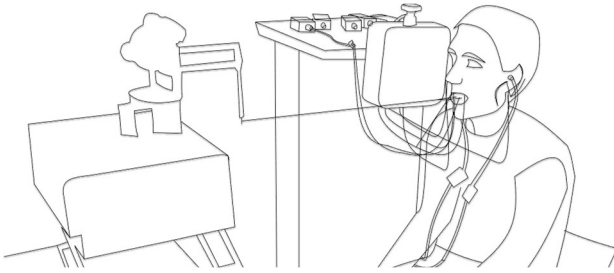


Figure 1: *Experimental setup*

The experimental setup is presented in Figure 1.

For the sensory perturbations, a tongue-stretch perturbation (PTB) and an altered auditory feedback perturbation (AAF) were used. For the tongue-stretch perturbation, we applied the same method as in our previous studies (Ito et al, 2020, Bourhis et al, submitted). A small robotic device (Phantom Premium 1.0, Geomagic) was connected to the tongue surface through a thin thread glued on both lateral sides of the tongue blade. A 1N force was applied in the forward direction as a step function with rise and fall phases of 5 ms, which prevents mechanical noise in the robot. For the auditory feedback perturbation, F1 was modified by 20 % either upward (incongruent with the effect of the mechanical perturbation) or downward (congruent with the effect of the mechanical perturbation) using Audapter (Cai et al., 2011). The altered sound was played back with 70 dB of white noise through magnetic compatible earphones (Natus Tip 300).

We recorded displacements of the tongue and jaw using electromagnetic articulography (Wave, Northern Digital Inc.). Six sensors were attached to the upper lip, lower lip, jaw, tongue tip, blade and dorsum in the mid-sagittal plane of the head. Reference sensors were also attached to the nasion, left and right mastoids, and the upper incisor for head movement correction. For each participant, the palate contour in the midsagittal plane was recorded by tracing the surface of the palate with a sensor glued on the experimenter's finger. The data were sampled at 200Hz. The produced speech sounds were also recorded using Audapter (Cai et al, 2011) at a 11.025kHz sampling rate: the first four formants, F1, F2, F3, and F4, were extracted at a sample frequency of around 345Hz.

In the test, the participants were asked to sustain vowel /ε/ for 3s in response to a visual cue. Vowel production started and ended with closed mouth position. Each trial was triggered manually by the experimenter after checking that the participant was ready. The two perturbations (PTB and AAF) were applied 1s after the onset of the vocalization. The tongue perturbation lasted for 1s. The auditory perturbation lasted until the end of the trial for a total duration of 4s. We tested five perturbed conditions combining auditory and tongue perturbations: altered auditory feedback alone (AAFup and AAFdown), tongue perturbation alone (PTB), and altered auditory feedback with tongue perturbation (AAFup+PTB and AAFdown+PTB). In total, 225 trials were carried out. The perturbation was applied in a pseudo randomly selected one third of the trials, so that the mechanical perturbation was never applied in two consecutive trials. Each of the five perturbed conditions was applied once within blocks of 15 trials. In total, 15 responses per condition were recorded (15 blocks).

We focused on the analysis of the acoustical data. Trials with wrong formant estimation ($F1 < 300$ Hz or $F1 > 700$ Hz) were removed from the analysis. Two participants were removed from the analysis due to high trial-to-trial variability. Acoustic data were aligned by the onset of the tongue perturbation, and

they were averaged across perturbed trials in each condition and in each participant. To remove individual variability of F1 amplitude, F1 was normalized by dividing it with the baseline amplitude that is the value averaged over the 50ms interval preceding the onset of the auditory perturbation.

We first compared AAFup+PTB and AAFdown+PTB and assessed whether auditory feedback changes the compensatory response to the tongue-stretch perturbation. As shown in our previous studies (Ito et al. 2020, Bourhis et al. submitted), the tongue stretch perturbation induces a decrease of F1 and the compensatory response reduces this decrease. We compared the peak amplitudes of the initial decrease and of the time course of the compensatory response. The times of these peaks were obtained based on the average response calculated over the three conditions involving tongue-stretch perturbation (PTB, AAFup+PTB and AAFdown+PTB). The time points of interest are labeled as P1 and P2 in Figure 2. The peak amplitudes were calculated over 20ms windows centered at these time points.

To characterize the role of the auditory feedback, we also compared AAFup and AAFdown conditions and assessed whether these two responses diverged or remained similar over the course of the vowel production. For this analysis, we focused on two time points, namely T_{base} : at the baseline and T_{diff} : at the onset time of the divergence between the two formant responses (Figure 2). T_{base} was set 150ms before the perturbation onset. To detect T_{diff} we applied a cluster-based analysis (Groppe et al, 2011). The procedure is based on a permutation test repeated 1000 times that was applied at each sampling point. We took in consideration the onset time of the first interval in which reliable difference was found over a set of consecutive sampling points (i.e a cluster). The amplitude at each time point was obtained using a 50ms window centered at this point.

A repeated measure ANOVA was applied in each amplitude comparison.

Results

We first compared the F1 responses to the tongue-stretch perturbation in two auditory conditions (AAFup+PTB and AAFdown+PTB). The normalized F1 responses in these two conditions are represented in the bottom panel of Figure 2. As in our previous studies (Ito et al, 2020, Bourhis et al, submitted), the tongue-stretch perturbation changed F1 and induced compensatory responses. The normalized F1 value decreased to about 0.83 116ms after the perturbation onset and the compensatory response brought it back to about 0.9 240ms after the perturbation onset. The temporal pattern of the responses is similar in two auditory conditions. Figure 3 represents the difference between the two auditory conditions in the normalized formant values measured at the times of the peak formant decrease (P1) and of the peak of the compensatory response (P2). We applied a two-way ANOVA on these formant values (time: P1 vs P2 and auditory condition: upshift vs downshift). There was no significant difference between auditory conditions ($p > 0.86$), but a significant difference exists in the time factor ($p < 0.001$). The interaction between the two factors was not significant ($p > 0.83$). These results indicate that a compensatory response was systematically induced by the tongue stretch-perturbation and that this response was not significantly affected by the auditory perturbation.

To verify the effect of AAF perturbation alone, we also compared the two auditory conditions without tongue-stretch perturbation (AAFup and AAFdown). The top panel of Figure 2 represents the normalized F1 response in these two

conditions. These two responses are similar in the time interval from the perturbation onset to time P2, and start diverging around 300ms. A cluster analysis reveals that this divergence starts from 360ms. Based on this analysis, we focused on the two time points T_{base} and T_{diff} shown in Figure 2.

The differences in normalized F1 value at these time points between the two auditory conditions in the absence of tongue stretch perturbation are represented in the left panel of Figure 3. A significant difference exists at T_{diff} , but not at T_{base} , indicating that F1 produced by the participants was significantly modified in response to its alteration induced by the auditory perturbation. As expected, this change was induced in a direction opposite to that of the perceived formant shift. The results also indicates that the compensation in response to the auditory perturbation was induced with a longer latency than the compensatory response to the tongue-stretch perturbation.

A divergence was also observed in the normalized F1 responses observed in the two auditory conditions in presence of the tongue-stretch perturbation (bottom panel of Figure 2). However, the difference at time T_{diff} , was not significant (see the right panel in Figure 4). The cluster based analysis also showed a difference occurring at a much later time (>1.2 s). This late difference detection may be due to the particularly large inter-participants variability associated with the tongue-stretch perturbation.

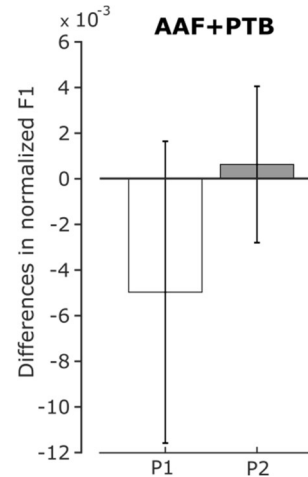


Figure 3: Differences in normalized F1 values between the two auditory conditions in presence of the tongue stretch perturbation (upshift: AAFup+PTB and downshift: AAFdown+PTB) at focused time points of interest (P1 and P2 in Figure 2). P1 corresponds the peak of the initial decrease of F1 and P2 corresponds the peak of the compensatory response. The error bars represent the standard error across participants. See methods for details.

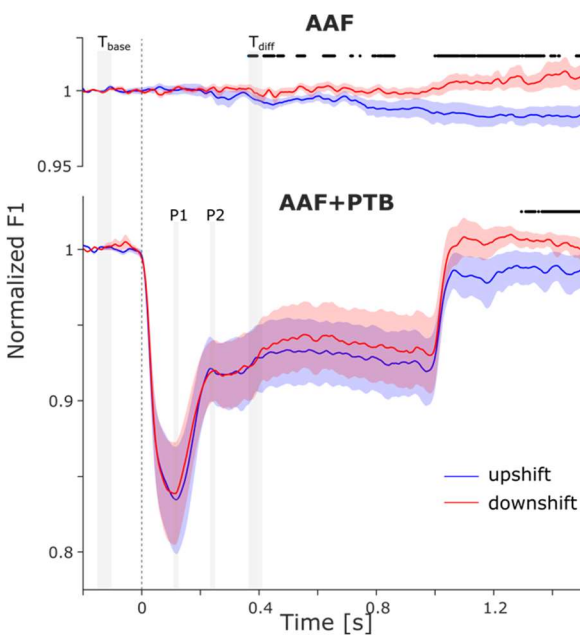


Figure 2: Normalized F1 responses in AAF (top panel) and AAF+PTB (bottom panel) conditions. The colored shaded areas represent the standard errors across participants. The vertical grey bars represent the times for which a comparison between auditory conditions was made. The black dots at the top of each panel represent the sample points at which a significant difference between the auditory conditions was revealed by the cluster-based analysis. See methods for details.

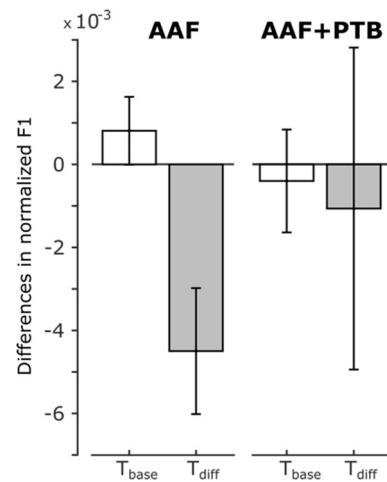


Figure 4: Differences in normalized F1 value between the two auditory conditions (upshift and downshift) in the (left panel) and presence (right panel) of the tongue stretch perturbation at two time points (T_{base} and T_{diff} in Figure 2). The error bars represent the standard error across participants. See methods for details.

Discussion and conclusion

In this study, we combined a sudden tongue-stretch perturbation with a shift of the first formant F1 during the steady-state production of vowel /e/. The formant shift was either in the same direction as the acoustic impact of the tongue perturbation or in the opposite direction.

The latency of auditory compensations for alterations of the auditory feedback involving formant shifts was shown (Cai et al., 2011) to be similar to the latency of the response to the tongue-stretch perturbation observed in our experiment, when

the auditory perturbation was applied during the production of a time-varying sequence of sounds (dynamical speech production henceforth). The decrease of the latency of auditory corrections in dynamical speech production, compared to static speech production, has also been observed in studies using pitch perturbation. The response latency was shorter when the pitch perturbation was applied during disyllabic sequences (100-150ms) (Donath et al., 2002; Xu et al., 2004) than the ones (150-200ms) when the perturbation was applied in the sustained vowel (Larson et al., 2000). Thus, the shortest latency of auditory-based corrections is comparable with the observed latency of the response to the tongue-stretch perturbation (around 130ms). Hence, in our experiment the auditory-based correction of the formant shift may influence the compensation for the tongue perturbation.

In line with previous studies (Ito et al 2020, Purcell et al, 2006), in our study both perturbations induced a compensatory response, and they were not simultaneous. The tongue-stretch perturbation induced a quick compensatory response with an average latency of 116ms. In contrast, the latency of auditory correction of the formant shift occurred significantly later, with an average latency of 360ms. Importantly the quick compensatory response to the tongue-stretch perturbation was not influenced by the additional formant shifts. These results confirm that the quick compensatory response relies on somatosensory feedback alone, and they suggest that somatosensory and auditory feedback control mechanisms may work separately and sequentially, due to their clearly different latencies.

Hence, our results show that, despite the tongue movement induced by the stretch perturbation during steady-state vowel production, the latency of the correction of the formant shift is the same as in usual steady-state vowel production (~ 400ms in Purcell et al., 2006). This suggests that auditory feedback is dependent on the planned speech production task (steady-state in our experiment) and not on whether or tongue movement occurs. We could expect different results if the tongue perturbation was applied during dynamical speech production. Just like the auditory-based corrections, the somatosensory response to perturbation could feature a significantly shorter latency.

In addition to the latency, the amplitudes of compensation were different between the two perturbations. The produced sounds were changed by about 15 % due to the tongue perturbation (see Figure 2) and 20 % due to the auditory perturbation. While the somatosensory-based compensation induced a recovery of around 50 % of the change induced by the tongue perturbation, the auditory compensation recovered only a few percent of the formant shift. Since in the tongue perturbation auditory change is associated with a compatible somatosensory error, it is easy to compensate for both sensory errors simultaneously. This is different from the case that the formant shift that is not associated with any somatosensory error.

Although the amplitude of compensation due to AAF perturbation was relatively small, this results still indicate the involvement of auditory error-detection mechanism. This detected error may be used in adaptation mechanism.

Overall, our results showed that somatosensory feedback induced faster compensatory responses than auditory feedback. For on-line speech motor control, these two compensatory mechanisms could be involved in different temporal phases, independently.

Acknowledgements

This work was supported by grants from the Agence Nationale de la Recherche (ANR-21-CE28-0022, PI. Takayuki Ito) and the National Institute on Deafness and Other Communication Disorders (Grant R01-DC017439).

References

- Bourhis, M., Perrier, P., Savariaux, C., Ito, T. (Submitted). Quick speech motor correction in the absence of auditory feedback.
- Cai, S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2011). Focal manipulations of formant trajectories reveal a role of auditory feedback in the online control of both within-syllable and between-syllable speech timing. *Journal of Neuroscience*, 31(45), 16483–16490. <https://doi.org/10.1523/JNEUROSCI.3653-11.2011>.
- Donath, T. M., Natke, U., & Kalveram, K. Th. (2002). Effects of frequency-shifted auditory feedback on voice F0 contours in syllables. *The Journal of the Acoustical Society of America*, 111(1), 357–366. <https://doi.org/10.1121/1.1424870>.
- Feng, Y., Gracco, V. L., & Max, L. (2011). Integration of auditory and somatosensory error signals in the neural control of speech movements. *Journal of neurophysiology*, 106(2), 667–679. <https://doi.org/10.1152/jn.00638.2010>.
- Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event - related brain potentials/fields I: A critical tutorial review. *Psychophysiology*, 48(12), 1711–1725. <https://doi.org/10.1111/j.1469-8986.2011.01273.x>.
- Ito, T., Szabados, A., Caillet, J. L., & Perrier, P. (2020). Quick compensatory mechanisms for tongue posture stabilization during speech production. *Journal of Neurophysiology*, 123(6), 2491–2503.
- Ito T, Bouguerra M, Bourhis M, Perrier P (2024) Tongue reflex for speech posture control. *Scientific Reports*, 14(1):6386.
- Larson CR, Burnett TA, Kiran S, Hain TC (2000) Effects of pitch-shift velocity on voice F0 responses. *J Acoust Soc Am* 107:559–564.
- Nasir, S. M., & Ostry, D. J. (2008). Speech motor learning in profoundly deaf adults. *Nature Neuroscience*, 11(10), 1217–1222. <https://doi.org/10.1038/nn.2193>.
- Patri, J. F., Ostry, D. J., Diard, J., Schwartz, J. L., Trudeau-Fisette, P., Savariaux, C., & Perrier, P. (2020). Speakers are able to categorize vowels based on tongue somatosensation. *Proceedings of the National Academy of Sciences*, 117(11), 6255–6263.
- Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J., ... & Zandipour, M. (2000). A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *Journal of Phonetics*, 28(3), 233–272.
- Purcell, D. W., & Munhall, K. G. (2006). Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *The Journal of the Acoustical Society of America*, 120(2), 966–977. <https://doi.org/10.1121/1.2217714>.
- Savariaux, C., Perrier, P., Orliaguet, J. P., & Schwartz, J. L. (1999). Compensation strategies for the perturbation of French [u] using a lip tube. II. Perceptual analysis. *The Journal of the Acoustical Society of America*, 106(1), 381–393.
- Tremblay S, Shiller DM, Ostry DJ (2003) Somatosensory basis of speech production. *Nature* 423:866–869.
- Xu, Y., Larson, C. R., Bauer, J. J., & Hain, T. C. (2004). Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences. *The Journal of the Acoustical Society of America*, 116(2), 1168–1178. <https://doi.org/10.1121/1.1763952>.