# THE ROLE OF FEEDBACK IN LIP-TUBE PERTURBATION OF TAIWAN MANDARIN ROUNDED VOWELS

Jian-zhi Huang[1] & Chenhao Chiu[1,2]

[1]Graduate Institute of Linguistics, National Taiwan University;
[2]Neurobiology and Cognitive Science Center, National Taiwan University
r10142007@ntu.edu.tw; chenhaochiu@ntu.edu.tw

## ABSTRACT

Previous lip-tube perturbation experiments have focused on the speaker's compensation capacity, but the role of feedback is less concerned. Previous studies have demonstrated that robust compensation may be induced from multisensory feedback. The present study explores whether somatosensory feedback alone can help with the compensation. The experiments involve lip-tube perturbation on Taiwan Mandarin rounded vowels /u/ and /y/ with auditory-somatosensory feedback and somatosensory feedback only. The acoustic results show that speakers compensate for the lip-tube perturbation in F2 and F3 in both feedback types. Tongue postures do not change significantly when either feedback type is provided in response to the perturbation, suggesting that tongue may not necessarily be a main articulator in contribution to the compensation but other articulators may be much more important. Generally, our results show that somatosensory feedback may play a fundamental role in compensation for lip-tube perturbation.

**Keywords**: feedback types, somatosensory feedback, lip-tube perturbation, internal model, lip-rounding

## 1. INTRODUCTION

In speech production, two major hypotheses have focused on two different aspects. One claims that speech targets are acoustically defined and regulated by auditory feedback (e.g., [1]). The other, e.g., Articulatory Phonology [2], posits that speech targets are articulatorily defined and composed of articulatory gestures. During speech production, feedback plays an important role, and the subsystems of the internal model [3]—inverse model and forward model—contribute to different aspects in speech motor control. The inverse model is associated with the muscular composition for a particular motor behaviour whereas the forward model is responsible for predictions, i.e., the expected outcomes of the planned action. The predicted outcome of the movement and the motor commands then will be checked to see if they correspond. If so, the motor commands are issued to the muscles. If there is a mismatch, then an error signal is generated to modify the motor commands. Therefore, sensory feedback used here is considered as closed loop, and sensory feedback monitors intended movements and reduces errors if any. Such a process is like an on-line monitoring system.

Lip-tube perturbation used a lip-tube with 2cm in diameter (to prevent lip-rounding of narrower lip areas) held between the lips with one end against the incisors to perturb lip-rounding. The function of the lip-tube was to create wider lip areas in which is not favourable to lip-rounding gestures. The experimental design aimed to observe whether speakers can still produce the target rounded vowels while lips are prohibited to create narrower lip areas and what kind of articulatory compensations (tongue movements) they made in respond to the lip-tube perturbation. Lip-tube perturbation in previous studies [4, 5] has focused on the compensation capacity and how speakers use the internal model to adapt to the lip-tube perturbation, but the role of feedback in guiding the speakers to use the internal model to reach speech targets has not been carefully examined in previous lip-tube perturbation studies. Several perturbation studies focusing on feedback have demonstrated that without the auditory feedback, somatosensory feedback alone can still constitute speech goal [6], and auditory feedback is not dominant—both auditory feedback and somatosensory feedback can be beneficial to the speakers [7]. However, it is yet to be determined that whether the observed results from previous lip-tube perturbation studies [4, 5] are consequences of reinforcement from multisensory feedback, i.e., speakers are accessible to multiple sensory feedback, so that they might have more accessible resources in which some are potentially beneficial to them. In this case, the auditory-somatosensory feedback received by the speakers in previous lip-tube perturbation studies might somehow bias the results which might simply be the consequence of changed acoustics. Thus, the present study aims to explore whether somatosensory feedback alone can help with the compensation.

## 2. METHODS

### 2.1. Participants

Participants were four native Taiwan Mandarin speakers (balanced gender, aged 22–27). The participants did not report any history of auditory abnormality or speech production disorder.

### 2.2. Stimuli and procedure

The stimuli used in the experiment were two Mandarin rounded vowels /u, y/ with four lexical tones, yielding eight monosyllabic words. The experimental procedure closely followed by [4, 5] with some modifications on the feedback types received by the speakers. Two feedback types were focused in the present study: auditory-somatosensory feedback and somatosensory feedback only. Each feedback type contained three conditions: baseline (10 repetitions), lip-tube perturbation (LP, 40 repetitions), and post-perturbation (PP, 10 repetitions). In lip-tube perturbation, a lip-tube of 2cm in diameter (to prevent lip-rounding of narrower lip areas) and 2cm in length (chosen to avoid lengthening the labial constriction) was held between the lips with one end against the incisors (see Fig. 1. left). During baseline and post-perturbation, no lip-tube was inserted.

The experiment, different from previous studies [4, 5], employed a within-subject design with auditory-somatosensory feedback and with somatosensory feedback only. When speakers receive auditory-somatosensory feedback, they can hear their own voice while at the same time feeling their own lips being perturbed and how they move their own tongues. When speakers receive somatosensory feedback only, their own voice was blocked and thus they were only accessible to the feelings of the perturbed lips and their own tongues. The order of receiving the two feedback types was counterbalanced. In somatosensory feedback only situation, two headphones were attached: the canal type in-ear headphone, playing background noise with *Pastoral Symphony,* and the bone conduction headphone, playing multi-talker babbling (see Fig. 1. right). The volume was adjusted to a level that the speakers could not hear their own voice. The participants were instructed to maintain the words pronounced as consistently and accurately as possible. To observe tongue postures and any tongue movements as articulatory compensations in response to the lip-tube perturbation, a portable ultrasound machine (CGM OPUS 5100) was employed to capture images in real time through a transvaginal electronic curved array probe (CLA 651). Speakers were instructed to sit upright wearing an ultrasound stabilization headset [8]. Acoustics and ultrasound images were recorded simultaneously throughout the experiment.

### 2.3. Data analysis

In terms of acoustic data, formants (F1, F2, and F3) were obtained at 20 duration-normalized timepoints through Praat[9] and were submitted to generalized additive mixed models (GAMMs) [10]. Ultrasound videos were first automatically traced by DeepEdge [11] for tongue contours and then were manually checked and corrected in GetContours [12]. Each tongue contour with 100 flesh points were then generated. Tongue contours were obtained from the midpoint of the vowel and were submitted to GAMMs [10]. Significant acoustic differences between lip-tube perturbation and baseline across the normalized time are defined as failed compensation. If speakers show successful compensation, there should be less significant differences.

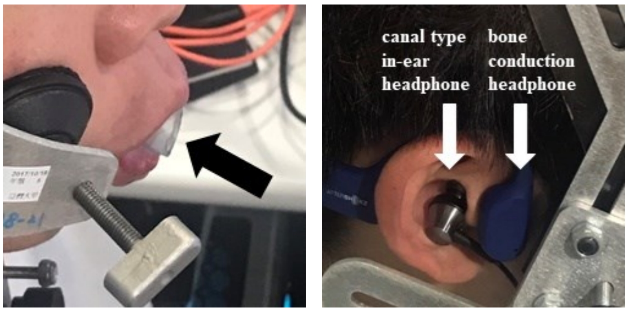


**Figure 1**: Experimental setup: lip-tube insertion (left) and headphones placement (right).

## 3. RESULTS

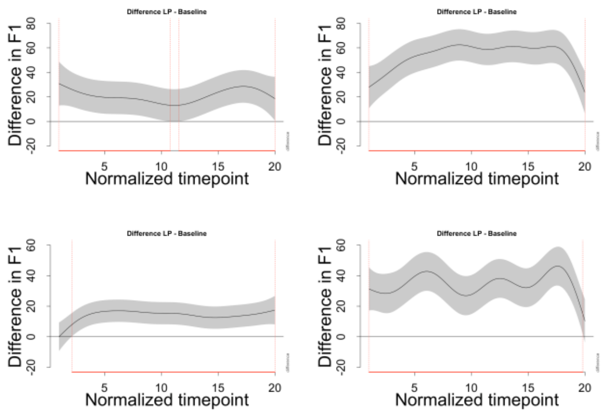


**Figure 2**: F1 differences of /u/ (upper) and /y/ (lower) between LP and Baseline when participants received auditory-somatosensory feedback (left) and somatosensory feedback only (right). Significant differences were enclosed by red dotted lines.

In both feedback types, significant differences of acoustics in terms of the three formants between post-perturbation (PP) and baseline were not found, suggesting that there is no hypercorrection or after-effect from perturbation. F1 differences between lip-tube perturbation (LP) and baseline for both rounded vowels with auditory-somatosensory feedback and somatosensory feedback only were all significantly different (Fig. 2.), suggesting that speakers cannot compensate for lip-tube perturbation in the dimension of F1 no matter which kinds of feedback are given.

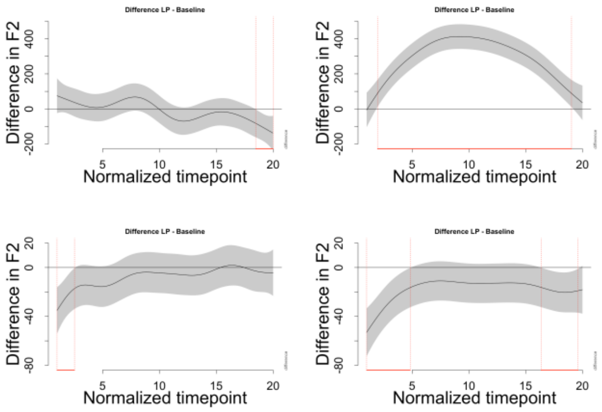


**Figure 3**: F2 differences of /u/ (upper) and /y/ (lower) between LP and Baseline when participants received auditory-somatosensory feedback (left) and somatosensory feedback only (right). Significant differences were enclosed by red dotted lines.

In terms of F2 differences, in the case of /u/, successful compensation was only found when participants were provided with auditory-somatosensory feedback (Fig. 3. upper left) but not with somatosensory feedback only (Fig. 3. upper right). Differences between LP and baseline revealed that F2 of /u/ produced with auditory-somatosensory feedback was more consistent in general (as evidenced by limited differences in Fig. 3. upper left) while F2 of /u/ produced with somatosensory feedback only showed significant differences enclosed by the red dotted lines (Fig. 3. upper right).

In the case of /y/, successful compensations were found in both auditory-somatosensory feedback (Fig. 3. lower right) and somatosensory feedback only (Fig. 3. lower left). In both cases, speakers tried to maintain F2 even when lip-tube was inserted, thus only a minority part of the trajectory (towards either one end or two) showed significant differences between LP and baseline.

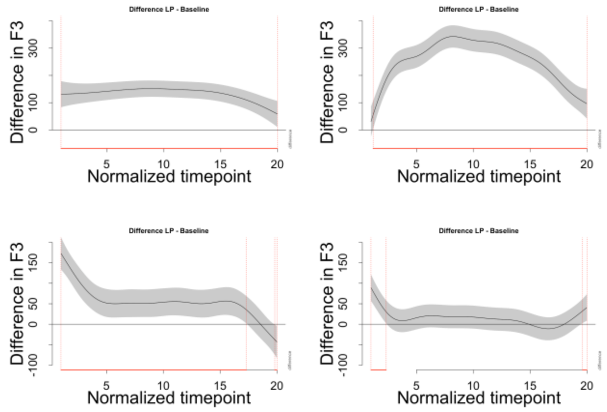


**Figure 4**: F3 differences of /u/ (upper) and /y/ (lower) between LP and Baseline when participants received auditory-somatosensory feedback (left) and somatosensory feedback only (right). Significant differences were enclosed by red dotted lines.

In terms of F3 differences, in the case of /u/, successful compensation was not found when both feedback types were given to the participants (Fig. 4. upper). F3 differences between lip-tube perturbation (LP) and baseline for /u/ with auditory-somatosensory feedback and somatosensory feedback only were all significantly different, showing that speakers cannot compensate for lip-tube perturbation for /u/ in the dimension of F3 no matter which kinds of feedback are given.

In the case of /y/, successful compensation was not found when receiving auditory-somatosensory feedback (Fig. 4. lower left) but when receiving somatosensory feedback only (Fig. 4. lower right). In Fig. 4. lower left, significant differences enclosed by rot dotted lines showed the majority part of the F3 trajectory was largely modulated. However, in Fig. 4. lower right, only very limited of the F3 trajectory was affected, suggesting the production remained intact.

For the articulatory results, tongue postures of /u/ no matter which kinds of feedback (Fig. 5. and 6.) are given did not change significantly in response to the perturbation (LP) or after the perturbation (PP). Also, the overall tongue postures when receiving auditory-somatosensory feedback (Fig. 5.) are similar to the tongue postures when receiving somatosensory feedback only (Fig. 6.).

Similarly, tongue postures of /y/ no matter which kinds of feedback (Fig. 7. and 8.) are given did not change significantly in response to the perturbation (LP) or after the perturbation (PP), and the overall tongue postures when receiving auditory-somatosensory feedback (Fig. 7.) are similar to the tongue postures when receiving somatosensory feedback only (Fig. 8.).

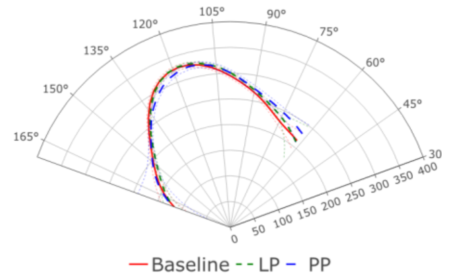


**Figure 5**: Tongue postures of /u/ when participants received auditory-somatosensory feedback. Tongue tip on the right. LP: lip-tube perturbation; PP: post-perturbation.

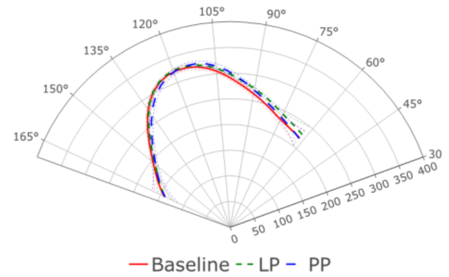


**Figure 6**: Tongue postures of /u/ when participants received somatosensory feedback only. Tongue tip on the right. LP: lip-tube perturbation; PP: post-perturbation.

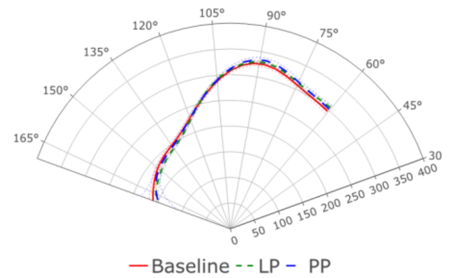


**Figure 7**: Tongue postures of /y/ when participants received auditory-somatosensory feedback. Tongue tip on the right. LP: lip-tube perturbation; PP: post-perturbation.

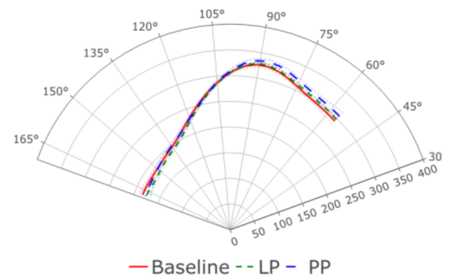


**Figure 8**: Tongue postures of /y/ when participants received somatosensory feedback only. Tongue tip on the right. LP: lip-tube perturbation; PP: post-perturbation.

## 4. DISCUSSION

First, we observe that the successful compensations are made in different formant dimensions for the two rounded vowels: F2 for /u/; F2 and F3 for /y/. This may suggest that the compensations made for the two rounded vowels focus on different dimensions.

Second, compensations can be successfully made not only with auditory-somatosensory feedback (F2 for both /u/ and /y/) but also with somatosensory feedback only (F2 and F3 for /y/). This echoes previous findings [6, 7] that somatosensory feedback alone can constitute speech goals. Also, we've found additional evidence for the claim that auditory feedback may not be dominant as in the case of F3 in /y/, where successful compensation was only found in the condition with somatosensory feedback only. This suggests that somatosensory feedback itself may be independent of auditory feedback and influential to perturbation compensation. In response to the two major hypotheses in speech production, the results show that somatosensory feedback alone (F2 and F3 for /y/) can still constitute speech goal and auditory feedback (F2 for /u/; F2 for /y/) might also play an equal role in constituting speech goal, and thus supporting both hypotheses.

Finally, no significant tongue movements were found across the tongue postures in response to the perturbation (LP) or after the perturbation (PP). Backward tongue movements reported in previous studies [4, 5] could be language-specific or due to the methods of dealing with articulatory data, where they all studied French and the articulatory data are not quantified and visualized as tongue postures but numerical transformations. This suggests that tongue may not necessarily be the main articulator in contribution to the compensation of lip-tube perturbation, but other articulators may be of more contribution. Another possible explanation is that lip-rounding may be organized by the coordination of different orofacial muscles which can be adjusted locally to reach the same global effect. Whether the lip-tube perturbation which constrains the lip areas is essentially related to the activation of the orofacial muscles would call for future studies.

## 5. CONCLUSION

The present study shows that somatosensory feedback itself can be independent of auditory feedback and it may play a fundamental role in compensation for lip-tube perturbation. Also, it does not necessarily mean that either auditory feedback or somatosensory feedback is more dominant, but rather they play in a parallel and relative relationship.

## 5. REFERENCES


[1] Jones, J. A., Munhall, K. G. 2000. Perceptual calibration of $F_0$ production: Evidence from feedback perturbation. *The Journal of the Acoustical Society of America*, *108*(3), 1246–1251.

[2] Browman, C. P., Goldstein, L. 1992. Articulatory phonology: An overview. *Phonetica*, *49*(3-4), 155–180.

[3] Wolpert, D. M., Ghahramani, Z., Jordan, M. I. 1995. An internal model for sensorimotor integration. *Science*, *269*(5232), 1880–1882.

[4] Savariaux, C., Perrier, P., Orliaguet, J. P. 1995. Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production. *The Journal of the Acoustical Society of America*, *98*(5), 2428–2442.

[5] Ménard, L., Perrier, P., Aubin, J. 2016. Compensation for a lip-tube perturbation in 4-year-olds: Articulatory, acoustic, and perceptual data analyzed in comparison with adults. *The Journal of the Acoustical Society of America*, *139*(5), 2514–2531.

[6] Tremblay, S., Shiller, D. M., Ostry, D. J. 2003. Somatosensory basis of speech production. *Nature*, *423*(6942), 866–869.

[7] Lametti, D. R., Nasir, S. M., Ostry, D. J. 2012. Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *Journal of Neuroscience*, *32*(27), 9351–9358.

[8] Articulate Instruments Ltd. 2008. *Ultrasound Stabilisation Headset User's Manual: Revision 1.4* (Articulate Instruments Ltd, Edinburgh, UK).

[9] Boersma, P., Weenink, D. 2021. Praat: doing phonetics by computer [Computer program]. Version 6.1.51.

[10] Wieling, M. 2018. Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics*, *70*, 86–116.

[11] Chen, W. R., Tiede, M., Whalen, D. H. 2020. DeepEdge: automatic ultrasound tongue contouring combining a deep neural network and an edge detection algorithm. Paper presented at the 12th International Seminar on Speech Production (ISSP 2020).

[12] Tiede, M., Whalen, D. H. 2015. Getcontours: An interactive tongue surface extraction tool. *Proceedings of Ultrafest VII*.