Speech adaptation to palatal perturbation: Evidence for sensorimotor reorganization across the workspace

Guillaume Barbier[1], Shari R. Baum[2], Lucie Ménard[3], Douglas M. Shiller[1]

Université de Montréal[1]; McGill University[2]; Université de Québec à Montréal[3]
barbier_guillaume@orange.fr; shari.baum@mcgill.ca; menard.lucie@uqam.ca; douglas.shiller@umontreal.ca

ABSTRACT

It has recently been shown that talkers are capable of simultaneously adapting multiple vowels to an alteration of auditory feedback during the production of complex, variable sentences [1]. The present study extends this work by examining adaptation to a more complex, physical alteration of the speech motor system (palatal prosthesis) that impacts both auditory and somatosensory feedback. Acoustic and kinematic measures (electromagnetic articulography of the tongue) were used to assess the initial impact of the perturbation on a range of vowels and consonants, as well as learned changes following 20 minutes of practice producing variable sentences with the prosthesis in place. Kinematic analyses revealed robust, systematic perturbation and motor learning effects across all speech sounds, indicating that talkers are capable of rapid adaptation in the production of multiple sounds across the articulatory workspace following a physical alteration of the vocal apparatus.

Keywords: Production, adaptation, oral kinematics

1. INTRODUCTION

In addition to a high degree of articulatory precision in speech production, competent talkers demonstrate considerable flexibility in oral motor patterns used to attain speech goals. The capacity for rapid, sensory-driven speech motor adaptation has been examined in numerous experimental studies involving real-time alterations of auditory feedback during speech production [1-7]**.** The vast majority of these studies have examined adaptation in very narrow contexts within any given experiment, e.g., one vowel within a small set of words [1-3], or a limited number of acoustically similar sentences [4-6]. In striking contrast to this prior work, a recent study has demonstrated that talkers are capable of reorganizing speech motor control across the entire vowel space following auditory feedback perturbations affecting multiple sounds during the production of variable sentences [7].

Sensorimotor adaptation to altered auditory feedback focuses on one component of sensory feedback during speech. However, speech has been shown to rely critically on both auditory and somatosensory feedback to achieve targets in both of these sensory domains [8]. In contrast to purely auditory manipulations, physical alterations of the speech apparatus involve changes in both auditory and somatosensory feedback, highlighting the role of both sources of sensory feedback in speech motor learning and control. Similar to studies involving altered auditory-feedback, studies examining adaptation to physical perturbations have, to date, been narrow in scope [9-12].

Here, we explore whether talkers are capable of adapting multiple speech sounds spanning the articulatory workspace (including various consonants and vowels) to a physical perturbation of the vocal tract involving a change in the shape of the hard palate, significantly impacting both auditory and somatosensory feedback. Palatal perturbations have been studied extensively in terms of acoustic changes related to a small number of sounds focusing on the alveolar/palatal region (e.g., /s/, /t/; [11,12]). While the acoustic effects appear limited to those sounds directly involving the palate, kinematic measures have suggested that this manipulation impacts articulatory speech movements beyond the palatal region [13].

Capitalizing on these prior findings, we used a palatal prosthesis combined with kinematic measures of the tongue to explore whether talkers are capable of rapidly adapting the control of multiple speech sounds across the articulatory workspace to a complex physical perturbation during production of complex, variable sentences.

2. METHODS

Nine adult participants (5 male, 4 female, native speakers of North-American English) with no history of speech, hearing or language disorder were tested. All procedures were approved by the IRB of the Faculty of Medicine, McGill University.

A custom-made palatal prosthesis was constructed for each participant by taking an alginate dental impression of the upper teeth and hard palate and producing a stone cast. A thermoplastic material was applied to the alveolar region of the cast to produce a rigid prosthesis following dimensions similar to those used in prior studies [11-13]: 6 mm thick ridge immediately behind to the upper incisors, tapering to 1 mm thickness at a distance of 2 cm.

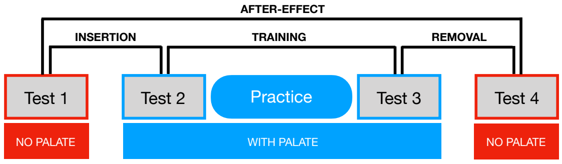
Participants carried out a series of four speech tests, each involving the production of 8 different symmetrical vowel-consonant-vowel pseudo-words containing the vowels /i/ and /æ/, combined with the consonants /s/, /t/, /k/ and /p/. Each word was produced 8 times in randomized order, totaling 64 productions per test. Four such speech tests were carried out in the following sequence (Figure 1): 1) immediately prior to insertion of the palatal prosthesis (*Test-1*), 2) immediately following insertion of the prosthesis and prior to the practice period with the palate in place (*Test-2*), 3) immediately following speech practice with the prosthesis in place (*Test-3*), and 4) immediately upon removal of the prosthesis (*Test-4*). Following speech *Test-2*, participants underwent a period of speech practice with the prosthesis in place, consisting of reading aloud a set of 47 sentences (drawn from the Harvard Sentences [14]) containing approximately balanced proportions of the four consonant sounds used in the speech test. The entire set of 47 sentences was read four times, each in a different randomized order (~20 minutes total duration).

Acoustic recording and analysis. Speech acoustics were recorded digitally (16-bit, 22500 kHz) using a directional microphone (ME66, Sennheiser, Germany) positioned 1 meter from the participant to reduce interference with the electromagnetic positional measurement system. For consonants, following prior studies of adaptation to palatal prostheses [12,13], analysis focused on the first spectral moment (spectral centroid) computed over a 20-msec window at the mid-point (for /s/) or aligned at burst onset (for /t/, /k/ and /p/). For vowels, F1 and F2 frequency were estimated using LPC over a 40-ms window centered at the vowel mid-point. The analyses of vowel formants focused only on the initial vowel in each VCV sequence (V_1), in order to minimize the contribution of carry-over co-articulation effects from the consonant.

For each subject, within each of the four speech tests, the consonant and vowel acoustic measures were first averaged across the 8 repetitions of each VCV context. These average measures were then transformed into a set of difference scores representing the four primary experimental effects of interest: 1) the effect of palate *insertion* (*Test-2* - *Test-1*), 2) the effect of speech *training* with the palate in place (*Test-3* - *Test-2*), the effect of palate *removal* (*Test-4* - *Test-3*), and 4) the *after-effect* (*Test-4* - *Test-1*; see Fig. 1, top).

Statistical analyses were subsequently carried out on these transformed scores, using repeated-measures ANOVA to test for effects of CONSONANT (s, t, k, p) and VOWEL (i and æ) contexts. One-sample t-tests (Holm-Bonferroni corrected) were used to test whether the transformed scores representing the effects of palate insertion, training, removal and after-effects were reliably different from zero.

Figure 1: The sequence of palate insertion (bottom), the different speech tasks (middle), and the four experimental effects of interest (top).



Kinematic measurement and analysis. Kinematic measurement of the tongue, jaw and head was carried out using an electromagnetic articulograph (AG500, Carstens, Germany), at 200 samples per second. Two sensors were placed mid-sagittally on the tongue, with one at the blade (1.5 cm from apex) and one at the dorsum (4.5 cm from apex), and one was affixed to the mandible at the lower incisors. Sensors affixed to the upper incisors, left mastoid and nasion were used to transform raw 3D tongue/jaw positions into a head-centered coordinate frame aligned with the occlusal plane (determined per subject using a triangular bite-plate with 3 sensors). The present analyses focused on tongue position in the mid-sagittal plane.

Kinematic data analysis followed the approach outlined for the acoustic measures. Tongue sensor positions (mid-sagittal x and y in mm), taken at the mid-point of production (for /s/, /i/ and /æ/) or burst onset (for /t/, /k/ and /p/), were averaged across the 8 repetitions of each VCV within each of the four speech tests. These averages were then converted into difference scores, yielding a vector (with an *amplitude* and *direction* to be analyzed separately) representing the tongue position change associated with the four key effects of interest: palate insertion, training, removal and learning after-effects.

Statistical analyses were carried out on measures of vector amplitude using rANOVA to test for effects of VCV CONTEXT (8 vowel/consonant combinations), and POSITION within the VCV (V_1, C, or V_2). One-sample t-tests were used to assess whether changes were reliably different from zero. For measures of vector directions, circular 1-way ANOVA (*Watson-Williams* test) was used to examine the difference between phonemes. Non-uniformity of angles among talkers was examined using *Rayleigh* tests [13].

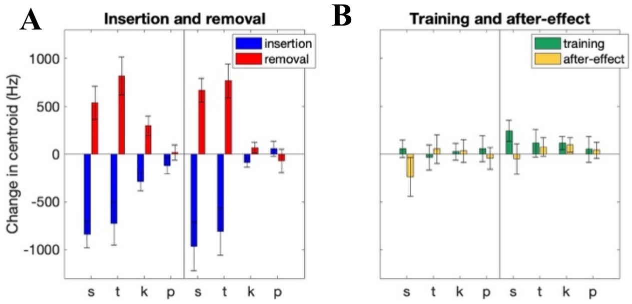
3. RESULTS

3.1 Acoustics – Consonants

Changes in consonant centroid frequency associated with palate *insertion* and *removal* are shown in Figure 2A. Immediately following palate insertion, centroid frequency is seen to decrease for production of /s/, /t/, and to a smaller degree, /k/ (blue bars). A change in the opposite direction is observed immediately upon palate removal (red bars).

For the insertion effect, a significant main effect of CONSONANT was found ($p < .05$), but not of VOWEL and no interaction. Averaging across vowel contexts, t-tests vs. zero showed a significant change following insertion for /s/, /t/, and /k/ ($p < .05$). The effect of palate removal mirrored the insertion effects, with a significant main effect of CONSONANT and significant t-tests vs. zero for /s/, /t/, and /k/ ($p < .05$).

Figure 2: Acoustic changes in consonant production. Error bars show ±1 SE.



Changes in centroid frequency following 20-minutes of speech *practice,* as well as the learning *after-effect*, are shown in Figure 2B and can be seen to be small in amplitude in comparison to the insertion/removal effects (for /s/ and /t/). No main effect of CONSONANT or VOWEL and no interaction were found ($p > .05$). The change vs. zero, averaged across all VCV contexts was also not significant. Because prior studies using palatal perturbations focused primarily on /s/ production, we directly examined whether this sound showed any effect following speech practice. The difference from zero for /s/ was found to be significant only in the low vowel context /æ/ ($p < .05$). Learning after-effects for /s/ in both vowel contexts were not significant.

3.2 Acoustics – Vowels

Figure 3A shows mean changes in F1 and F2 during vowel production (averaged over the 4 consonant contexts), associated with palate *insertion* and *removal*. Changes can be seen principally in F2, with limited effects in F1.

For F1, the main effects and interaction of VOWEL and CONSONANT were not statistically reliable, and no reliable difference from zero was observed for either palate insertion or removal. Similarly, for F2, none of the main effects or interactions were significant for palate insertion or removal. However, the change in F2 vs. zero (averaging across vowel and consonant contexts) was found to be statistically reliable for both insertion ($p < .05$) and removal ($p < .05$).

Figure 3: Acoustic changes in F1 (left half of each panel) and F2 (right half). Error bars show ±1 SE.

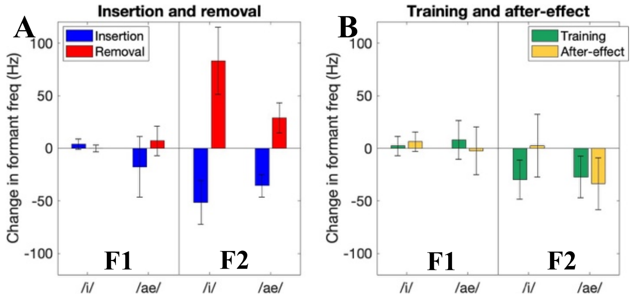
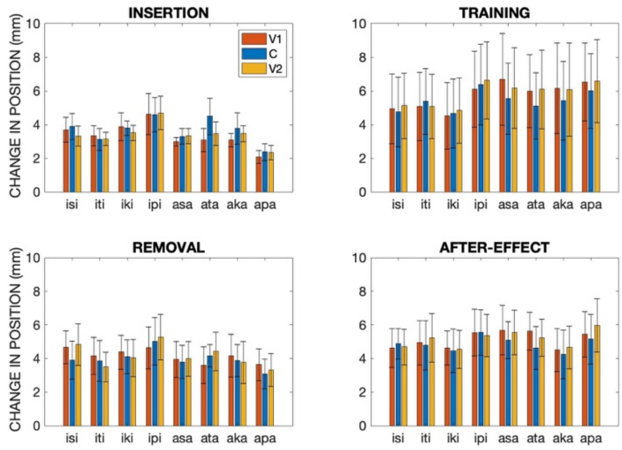


Figure 3B shows changes associated with *training* and the learning *after-effect* for F1 and F2. For F1, rANOVA main effects of VOWEL and CONSONANT , and the interaction, were not statistically reliable. No reliable difference from zero was observed for either insertion or removal. For F2, none of the main effects or interactions were significant. The change in F2 vs. zero (averaging across vowel and consonant contexts) was statistically reliable for training ($p < .05$), but not the after-effect.

3.3 Kinematics – Consonants and Vowels

The mean amplitude of the tongue positional change associated with each of the four experimental effects of interest is shown in Figure 4 for all consonants and vowels in each of the 8 VCV contexts (tongue blade sensor shown). A change in tongue position can be observed for all sounds, in all contexts.

Figure 4: Mean amplitude of changes in tongue blade sensor position. Error bars show ±1 SE.

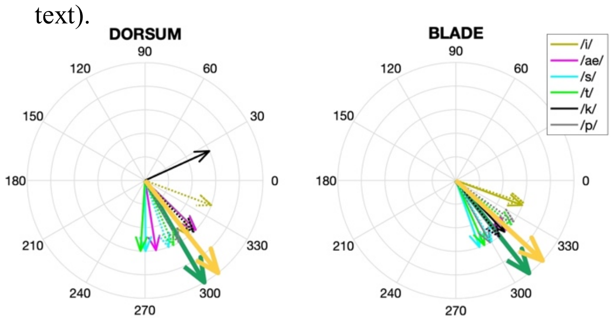


For the tongue blade sensor, the reliability of differences between the 8 VCV contexts and between the three VCV positions (V_1, C, V_2) was examined using rANOVA separately for each of the four experimental effects. None of the main effects or interaction effects were statistically reliable, reflecting the similarity in amplitude across contexts. The same pattern of main and interaction effects was observed in the tongue dorsum sensor.

The magnitude of tongue position changes for each of the experimental effects was compared to zero using single-sample t-tests, averaging across VCV contexts and VCV positions (consistent with the ANOVA results). For both tongue sensors, all four tests showed robust changes (insertion, removal and after-effect: $p < .01$; training: $p < .05$).

If the observed changes in tongue position associated with speech training truly reflect changes in feed-forward control (i.e., motor learning), we would predict that the *direction* of tongue positional changes associated with the after-effects (tongue position changes relative to baseline that persist following removal of the palatal prosthesis) would be consistent with the directions of the training effects. Figure 5 shows the average direction of tongue position change associated with speech training (small solid vectors) and after-effect (small dotted vectors), for each sound. Angles represent directions in the midsagittal plane, with 0° corresponding to the anterior direction. Grand average angles are shown as large vectors, with training in green and after-effect in yellow. A similar directional change in tongue position is seen for both effects, characterized overall by tongue advancement and lowering.

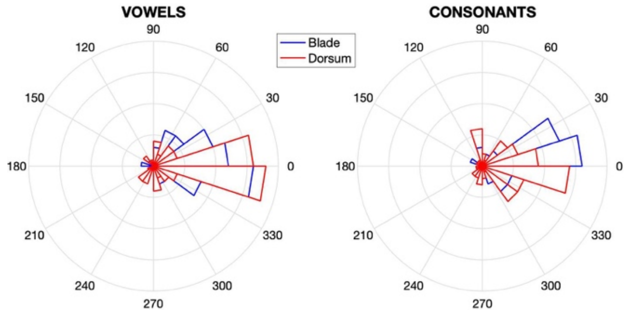
Figure 5: Directions of tongue positional change associated with the training and after-effect (see text).



The reliability of angle differences among the 6 phonemes was examined separately for each experimental effect. No reliable difference between sounds was observed for either sensor ($p > .05$). Further, no reliable differences were observed *between* training and after-effects for any of the 6 sounds ($p > .05$). *Rayleigh* tests examined the directional variation of training effects among the participants. No significant departures from circular uniformity ($p > .05$) were observed for any sound, confirming the variable nature of motor strategies among the different participants.

If tongue positional changes following speech training truly reflect motor learning, they should be directionally related to the observed after-effects irrespective of inter-speaker differences in the training patterns. This relationship was evaluated across all contexts by calculating the angular *difference* between the two effects separately for each of the VCV contexts and syllable positions within each subject. Figure 6 shows the distribution of these angle differences related to the production of vowels (left) and consonants (right), revealing that the individual training and after-effects are indeed characterized by a common mean directional difference close to 0° (mean angle difference for vowels: 10.3° for the tongue blade and -1.4° and for tongue dorsum; for consonants: 7.9° for the tongue blade and -1.2° for tongue dorsum).

Figure 6: Distribution of angular differences between the training and learning after-effects.



4. DISCUSSION

Acoustically, as shown in prior studies, reliable perturbation and training effects were observed in the production of /s/. Here, this finding was extended by the observation of reliable acoustic perturbation effects in the production of the velar /k/ as well as in the vowels /i/ and /æ/, with vowel production also showing a small but reliable training effect in F2. These results highlight the non-uniform effect of the palatal perturbation on speech acoustics, while also hinting at training effects beyond sounds involving direct interaction with the physically altered palate.

Kinematically, the amplitude of tongue positional changes revealed robust perturbation effects across all VCV contexts and positions, and subsequent training and learning after-effects in all cases. Strikingly, a comparable magnitude of tongue position change was observed between the various vowels and consonants, despite differences in the presumed interaction with the palate, and also despite observed differences in the acoustic effects among sounds. Directional analyses of tongue positional changes related to training and after-effects support the idea that the observed kinematic changes found across vowels and consonants indeed reflect learned changes in the feed-forward planning of speech movements.

The present acoustic and kinematic findings build upon the prior result of Lametti et al. [6] in demonstrating that talkers are capable of simultaneously adapting speech sounds across the articulatory workspace to a complex physical, multisensory perturbation.

7. REFERENCES

[1] Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. Science, 279(5354), 1213-1216.

[2] 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.

[3] Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. The Journal of the Acoustical Society of America, 122(4), 2306-2319.

[4] Cai, S., Ghosh, S.S., Guenther, F.H., and Perkell, J.S. (2011). Focal manip- ulations of formant trajectories reveal a role of auditory feedback in the on- line control of both within-syllable and between-syllable speech timing. J. Neurosci. 31, 16483–16490.

[5] Patel, R., Reilly, K.J., Archibald, E., Cai, S., and Guenther, F.H. (2015). Responses to Intensity-Shifted Auditory Feedback During Running Speech. J. Speech Lang. Hear. Res. 58, 1687–1694.

[6] Lametti, D.R., Smith, H.J., Freidin, P.F., and Watkins, K.E. (2018). Cortico-cerebellar Networks Drive Sensorimotor Learning in Speech. J. Cogn. Neurosci. 30, 540–551.

[7] Lametti, D. R., Smith, H. J., Watkins, K. E., & Shiller, D. M. (2018). Robust Sensorimotor Learning during Variable Sentence-Level Speech. Current Biology, 28(19), 3106-3113.

[8] Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. Nature, 423(6942), 866.

[9] Savariaux, Perrier, & Orliaguet. (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.

[10] Jones, J. A., & Munhall, K. G. (2003). Learning to produce speech with an altered vocal tract: the role of auditory feedback. The Journal of the Acoustical Society of America, 113(1), 532–543.

[11] Hamlet, S., Stone, M., & McCarty, T. (1978). Conditioning prostheses viewed from the standpoint of speech adaptation. Journal of Prosthetic Dentistry, 40(1), 60-66.

[12] Baum, S. R., & McFarland, D. H. (1997). The development of speech adaptation to an artificial palate. The Journal of the Acoustical Society of America, 102(4), 2353-2359.

[13] Thibeault, M., Ménard, L., Baum, S. R., Richard, G., & McFarland, D. H. (2011). Articulatory and acoustic adaptation to palatal perturbation. The Journal of the Acoustical Society of America, 129(4), 2112-2120.

[14] IEEE Recommended Practice for Speech Quality Measurements (1969). IEEE Trans. Audio Electroacoust. 17, 227–246.

[15] Berens, P. (2009). CircStat: a MATLAB toolbox for circular statistics. J Stat Softw, 31(10), 1-21.