

# Speaking with a Bite-Block: Data and Modeling

Majid Zandipour<sup>1,2\*</sup>, Joseph Perkell<sup>1,2\*</sup>, Frank Guenther<sup>2,1</sup>,

Mark Tiede<sup>1,3\*</sup>, Kiyoshi Honda<sup>4</sup>, Emi Murano<sup>5</sup>

<sup>1</sup>Massachusetts Institute of Technology, Speech Communication Group – R.L.E.  
50 Vassar Street, Cambridge, MA 02139 USA

<sup>2</sup>Boston University, Department of Cognitive and Neural Systems  
677 Beacon Street, Boston, MA 02215 USA

<sup>3</sup>Haskins Laboratories  
300 George Street, New Haven, CT 06511 USA

<sup>4</sup>LPP-Paris3  
19 rue des Bernardins, 75005 Paris France

<sup>5</sup>University of Maryland Dental School  
650 W Baltimore Street, Baltimore, MD 21209 USA

majidz@speech.mit.edu, perkell@speech.mit.edu, guenther@cns.bu.edu,  
tiede@haskins.yale.edu, honda@atr.jp, emurano@umaryland.edu

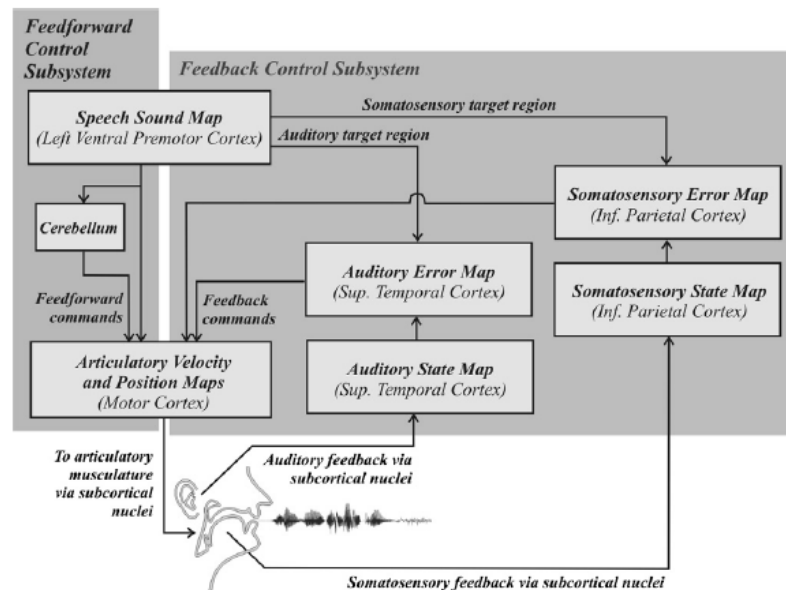
**Abstract.** *Based on the assumption that the goals of phonemic speech movements are both auditory and somatosensory in nature, a biomechanical model of the vocal tract in conjunction with an adaptive controller inspired by the DIVA model of speech production (Guenther et al., 2006) were employed to investigate how the feedforward and feedback mechanisms operate in the planning for the production of the vowel [iy] with the presence of a bite-block and/or masking noise. Simulations of the vocal tract model corroborate the hypotheses predicted by the DIVA model, suggesting that: a) in the absence of auditory feedback, the subject used his learned feedforward motor command for producing a vowel-to-vowel sequence; b) in the presence of a bite-block and auditory feedback, the subject relied on the acoustic feedback mechanism during his productions; c) in the absence of auditory feedback and with a bite-block present, the subject relied on his feedforward motor plans initially, and then later on, incorporated corrective commands from the somatosensory feedback system into his productions.*

---

\* Supported by NIH-NIDCD #R01-DC001925

## 1. Introduction

The data and simulations of the biomechanical vocal tract model described below show how feedforward and feedback mechanisms operate in the planning of phonemic articulatory movements in speech production. Based on the assumption that the goals of phonemic speech movements are both auditory and somatosensory (e.g., muscle lengths and articulator contact), a biomechanical model of the vocal tract in conjunction with an adaptive controller inspired by the DIVA model of speech production (Guenther et al., 2006) were employed to investigate how speech movements are planned with the presence of a bite-block and/or masking noise.



**Figure 1.** Schematic diagram of The DIVA model of speech motor planning (Guenther et al., 2006).

The DIVA model (Figure 1; from Guenther et al., 2006) hypothesizes that speech production is controlled by two parallel subsystems: a *feedback system* and a *feedforward system*, which are engaged by the activation of a *speech sound map*. The feedback system encodes sensory expectations, or goals, for the planned speech sound which are compared to the resulting auditory and somatosensory sensations. If there is a mismatch, it results in feedback-based corrections, which are incorporated into feedforward commands for future movements. Thus feedforward commands are learned and refined by incorporating corrective commands from the feedback controller. As speech acquisition progresses and the feedforward commands improve, feedback control is replaced by feedforward control for normal, fluent speech production. The DIVA model's feedforward and feedback control systems, when tuned to a particular vocal tract, encode information about relations between auditory and articulatory parameters.

To investigate the role of sensory feedback in speech production, articulatory and auditory parameters were measured while a subject clenched a bite-block between the teeth with and without masking noise. It has been shown previously that with the presence of auditory feedback and a bite-block, speakers were easily able to produce relatively accurate formants (Lindblom et al. 1978). However, Baum et al. (1996)

showed such compensation was not immediate; there were small but significant improvements in the acoustic characteristics of vowels that were produced subsequent to a period of conversation with a bite-block. Campbell (1999) investigated the role of auditory feedback with and without the presence of a bite-block and 80dB of masking noise, and showed that normal hearing subjects compensated substantially, but not completely, leading to the inference that auditory feedback played an important role in the compensation for the bite-block. Ito et al. (2006) applied unexpected perturbations to the jaw during vowel production, and showed that after initial deviation from control trajectories, F1 compensation began about 60ms after the onset of perturbation and was nearly complete after 200ms, even though perturbed jaw trajectories remained entirely distinct from baseline. The present study extends such work by using a biomechanical simulation of the vocal tract in conjunction with an adaptive controller to examine the role of sensory feedback and how feedforward and feedback mechanisms operate in the planning for the production of the vowel [iy] in the presence of a bite-block and/or masking noise.

### **1.1. Hypotheses**

The following hypotheses are predictions derived from the DIVA model for situations in which jaw movements are fixed with or without masking noise:

1. With masking noise and no bite-block, speakers will rely on feedforward commands for articulatory movements, resulting in movement and formant trajectories that are similar to their productions without masking noise.
2. With both masking noise and a bite-block, speakers will initially rely on feedforward commands for producing a vowel-to-vowel (V-V) movement because of the perturbed somatosensory feedback and lack of auditory feedback. Formant and movement trajectories will be significantly different from their normal productions. With practice, corrective commands issued by the somatosensory feedback system (based on somatosensory error signals) will be incorporated into feedforward commands, resulting in formant trajectories that are closer to those produced without perturbations.
3. With a bite-block and normal auditory feedback, speakers will rely on auditory feedback, producing movement and formant trajectories that approach normal.

## **2. Methods**

A male speaker of American English participated in two experiments. In the first experiment, the elicitation set consisted of  $V_1V_2V_1$  sequences recorded during normal and bite-block speaking conditions. For ten repetitions of [aa]-[iy]-[aa], the subject clenched on a 25mm plastic bite-block with his left premolars, while the speech, kinematic (EMMA), and EMG signals were recorded. In the second experiment, the elicitation set consisted of 20 repetitions of five hVd words. Speech and kinematic data were acquired in four speaking conditions in the following order: 1) normal, 2) with 95dB of masking noise, 3) with masking noise and a bite-block, 4) with the bite-block and normal auditory feedback. Before each utterance in the bite-block condition, a bite-block was inserted between the subject's upper and lower left premolars. The thickness

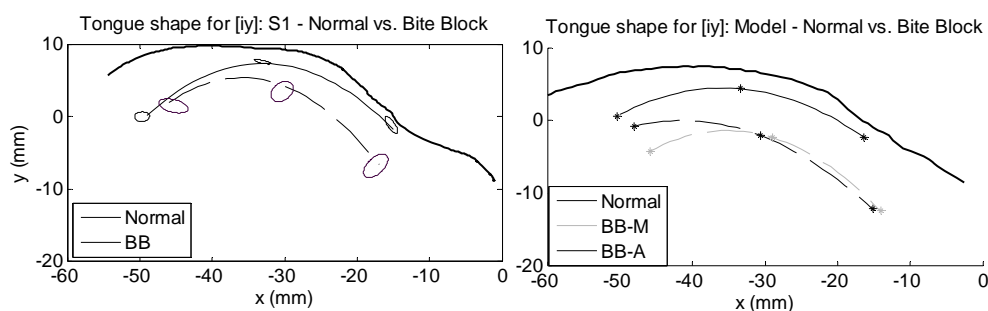
of the bite-blocks varied inversely with typical jaw opening for each vowel, as follows: for [iy] and [ih] 25 mm (to force a large jaw opening); for [aa] 5mm (to prevent normal jaw opening), and for [eh] and [ae] both sizes of bite-blocks were chosen randomly.

The subject's data from both experiments were simulated using a DIVA-like adaptive controller coupled to a vocal tract model (Zandipour, 2006) that included muscular anatomy, a model of muscle contraction (Cheng et al., 2000), surface contact and dynamical properties. The configuration of the vocal-tract model was adapted to the subject's vocal tract (using static, lateral MR images of the subject) and its output was compared with his data. The acoustic output of the vocal tract model is represented in mel space by a 64 point spectrum.

### 3. Results

#### 3.1. Experiment 1

The insertion of a 25 mm bite-block perturbed the production of [iy] in the [aa]-[iy]-[aa] utterance. The left panel in Figure 2 shows the subject's data (S1), in which the tongue shape (defined by a spline fit to three EMMA tongue transducers) for the normal (solid line) and bite-block (dashed line) productions of [iy] are depicted. In this study normal auditory feedback was available for both conditions. The ellipses show 95% confidence intervals for 10 repetitions. Note that the tongue in the bite-block condition was more anterior and positioned lower (more at the tongue tip and less at the tongue dorsum) relative to the normal condition. In order to understand the subject's bite-block data, two [aa]-[iy] sequences were simulated while keeping the jaw opening constant at the [aa] position. In one simulation, the model planned an [aa]-[iy] sequence using a motor space for planning the speech movements; i.e., planning was carried out by interpolating between the initial muscle configuration and a target muscle configuration for [iy]. In the second simulation, the model used an acoustic/auditory planning space,



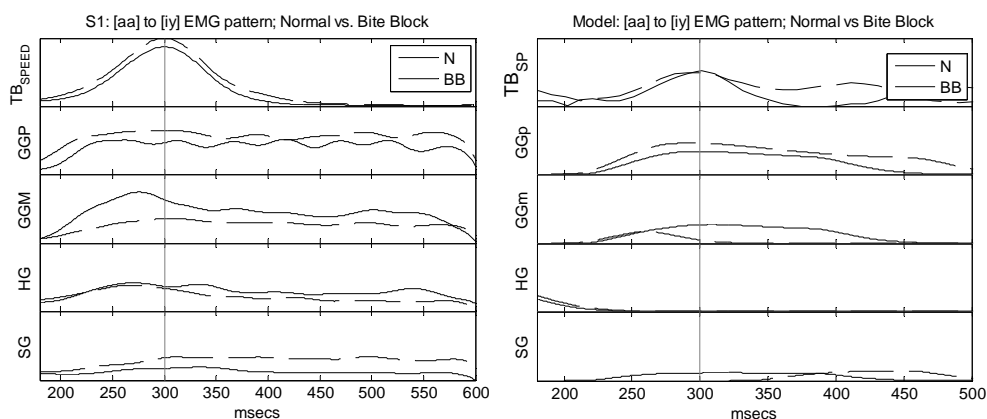
**Figure 2:** The left panel shows the data from the subject (S1) in which the tongue shape for the normal (solid) and bite block (dashed) productions of [iy] are depicted. The ellipses illustrate 95% confidence intervals. The right panel shows the shape of the tongue at [iy] generated by the model in normal (solid), bite block using feedforward motor plan (gray; BB-M), and bite block using acoustic feedback (dashed; BB-A) conditions. Locations of the tongue transducers are marked by stars (★). TT, TB, and TD are the location of the tongue tip, tongue body, and tongue dorsum EMMA transducers respectively.

interpolating between the current acoustic state and a target acoustic state for [iy]. The right panel in Figure 2 shows the shape of the tongue at [iy] generated by the model in

normal (solid), bite-block using motor planning (dash gray), and bite-block using acoustic planning (dash) conditions. Locations of the tongue transducers are marked by stars (\*). Note that the tongue in the bite-block condition using acoustic planning (BB-A) is lower and more anterior in relation to the normal condition, similar to the subject's data (left, BB). In addition, the simulated tongue shapes for the normal and bite-block using acoustic planning (BB-A) conditions show smaller displacement in the tongue dorsum (TD) than the tongue tip (TT), also as observed in subject's data. On the other hand, the simulated tongue position in the bite-block condition using motor planning (BB-M) is lower and parallel to the tongue in the normal condition, unlike the data.

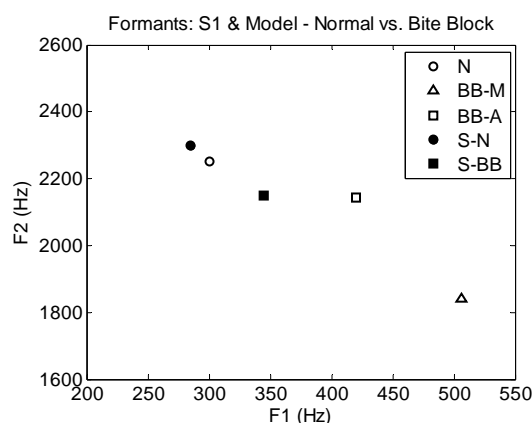
The right panel in Figure 3 shows the muscle activation signals (corresponding to EMG signals) from the vocal-tract model simulating an [aa]-[iy] transition in the normal speaking condition (solid lines) and with a bite-block using acoustic space planning (dashed lines). Similar to the subject's data (Figure 3, left panel), the model increased the activation of the Genioglossus Posterior (GGp) and Styloglossus (SG) slightly, and decreased the activation of the Genioglossus Medial (GGm) for the bite-block condition relative to normal production.

Figure 4 shows in the F1-F2 plane, the subject's data for [iy] in normal (●) and bite-block (■), and the model's simulation of the normal (○), bite-block using the motor planning scheme (△) and bite-block condition using the acoustic planning scheme (□) conditions. Note that the discrepancy between the subject's normal and the model's normal is due to discretization introduced by binning of the acoustic spectrum in the model. The formant values in the bite-block condition for the subject and acoustic planning scheme are the same along F2 and slightly different along F1, on the other hand, the formants for the subject and motor planning are very different along both



axes.

**Figure 3:** The right panel shows speed of the TB transducer ( $TB_{SP}$ ) and EMG signals from the vocal tract model simulating an [aa]-[iy] transition in the normal speaking condition (solid line) and with a bite block (dashed line). The vertical lines mark the peak speed of the tongue body transducer and the point of lining up the normal and BB conditions. Similar to the subject's data (left panel), the model increased the activation of the GGp and SG slightly, and decreased the activation of the GGm for the bite block condition relative to normal production.



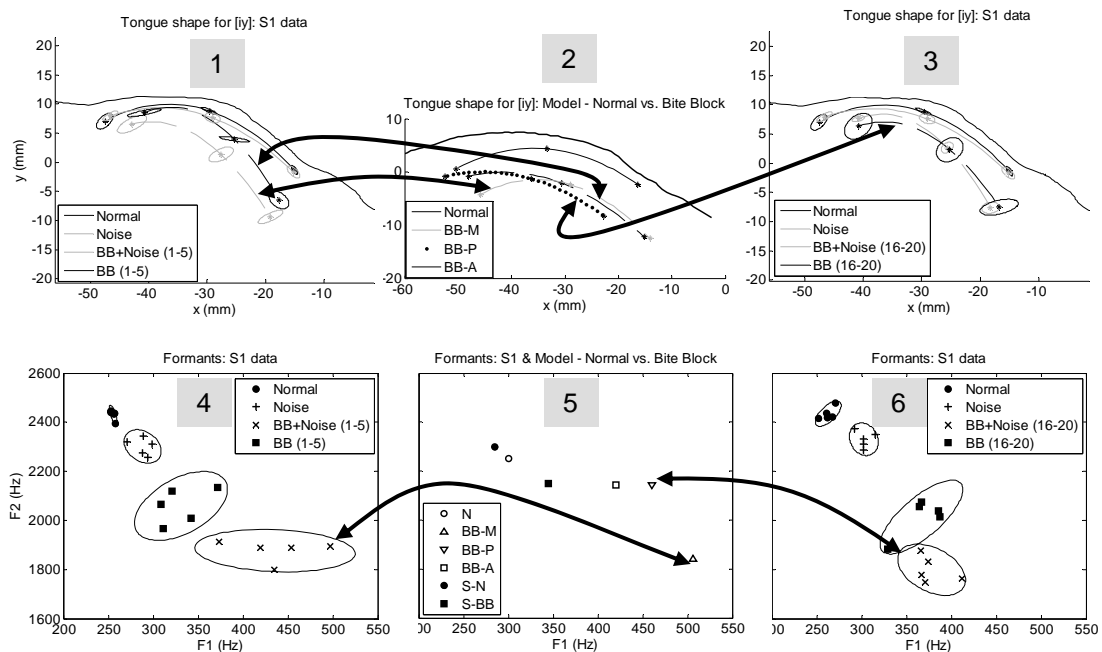
**Figure 4:** The subject's data for [iy] in normal (●) and bite block (■), and the model's simulation of the normal (○), bite block using the motor planning scheme (△, BB-M) and bite block condition using the acoustic planning scheme (□, BB-A) are shown in the F1-F2 plane.

These simulation and experimental results provide preliminary support for the use of an acoustic space over a motor space for articulatory movement planning.

### 3.2. Experiment 2

Panels 1 and 3 in Figure 5 show the subject's tongue shape for the first and last five repetitions of the word *heed* during the four experimental conditions: normal (black solid lines), noise only (gray solid lines), noise plus bite-block (gray dashed line), and bite-block only (black dashed line). As anticipated (Hypothesis 1), the production of [iy] with masking noise was similar to the normal production of that phoneme, suggesting that the subject relied on feedforward commands for articulatory movements. The initial productions during the bite-block only condition (panel 1, black dashed line) were better (smaller constriction between the tongue and palate, and closer formants to normal) relative to his initial production during the bite-block plus noise condition (panel 1, gray dashed line). Plots in panel 3 show the last five productions of [iy] in each speaking condition. Again, there was no difference between the normal and noise-only conditions, nor was there any difference between the last five and first five repetitions. The last five repetitions between the bite-block plus noise (panel 3, gray dashed line) and bite-block only (panel 3, black dashed line) conditions overlapped and were statistically indistinguishable. Comparison of the first (panel 1, gray dashed line) and last five repetitions (panel 3, gray dashed line) for the bite-block plus noise condition showed a shift in production, which according to the model, would have to be due to the use of somatosensory feedback.

Panel 2 shows the simulations of the model in the following speaking conditions: normal (black solid line), bite-block with auditory feedback (black dashed line), bite-block without any sensory feedback (gray dashed line), and bite-block with somatosensory feedback (gray dotted line). These simulations showed that the production of [iy] with a bite-block alone (black dashed line) was better (less constriction between the tongue and palate for TD, and formants closer to normal condition) than with no sensory feedback (gray dashed line). In addition, the model showed that producing an [iy] with a bite-block using somatosensory feedback (gray dotted line) would provide a better result than the condition without feedback (gray



**Figure 5:** Panels 1 and 3 show the subject's tongue shape for the first and last five repetitions of the word *heed* during the normal, noise only, noise plus bite block (panel 1, reps 1-5; panel 3, reps 16-20), and bite block only (panel 1, reps 1-5; panel 3, reps 16-20) conditions (see legends in each plot). Panel 2 shows simulations in the normal, bite block with auditory feedback, bite block without sensory feedback, and bite block with proprioceptive feedback conditions (see legends). The respective formant values are shown in panels 4-6. The black arrows show the correspondence between plots.

dashed line). The correspondence between the model's predictions and data is shown by the grey arrows between the panels in Figure 5.

Panels 4 and 6 present the acoustic results of the first and last 5 repetitions of [iy] in the F1-F2 plane during the four conditions respectively. Results in panel 4 show that the formant values during the bite-block plus noise condition (✕) were further from normal values (●) than were values during the bite-block only condition (■). Results in panel 6 show that after somatosensory feedback had been employed to compensate for the jaw perturbation, during the last five repetitions, F1 shifted significantly in the bite-block plus noise condition (✕) toward the normal production, whereas it shifted away slightly for the bite-block only (■).

Panel 5 presents the model's results in the F1-F2 plane, and shows that with respect to the normal values, there was a greater shift in formants for the bite-block without feedback (△) relative to the bite-block with auditory feedback (□), similar to the subject's data. The model showed that the production of an [iy] with bite-block and somatosensory feedback (▽) would result in a shift of the formants toward normal production relative to the bite-block without feedback (△), but not as much as the bite-block with auditory feedback (□).

These simulation and experimental results suggest that with erroneous somatosensory feedback and no auditory feedback, feedforward motor commands are relied upon initially and are modified with practice by corrections based on errors between expected and produced somatosensory feedback.

#### **4. Discussion**

To study the effects of jaw perturbation on the EMG, kinematics and formants, the subject's data for the production of the vowel [iy] with a bite-block were examined and simulated by the model successfully, and supported the following interpretations:

- In the absence of auditory feedback (noise only condition), the subject used his learned feedforward motor command for producing a vowel-to-vowel sequence.
- In the presence of a bite-block and auditory feedback, the subject relied on the acoustic feedback mechanism during his speech production.
- In the absence of auditory feedback and with the presence of a bite-block, the subject relied on his feedforward motor plans initially, and then later on, incorporated corrections from the somatosensory feedback system in his production.

Simulations of the vocal tract model corroborate the hypotheses predicted by the DIVA model, which posits a feedforward system and a sensory feedback system that operate in parallel during speech production.

#### **References**

- Baum, S. R., McFarland, D. H., & Diab, M. Compensation to articulatory perturbation: Perceptual data. *The Journal of the Acoustical Society of America*, 99(6):3791-3794, 1996.
- Campbell, M. *Articulatory compensation for a biteblock, with and without auditory feedback in hearing and hearing-impaired speakers*. Doctoral dissertation, City University of New York, 1999.
- Guenther, F.H., Ghosh, S. S., & Tourville, J. A. Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, 96(3):280-301, 2006.
- Ito, T., Tiede, M., & Ostry, D. A novel technique for triggering mechanical perturbations based on formant trajectories. In *Proceedings of the 5<sup>th</sup> International Conference on Speech Motor Control (Nijmegen)*, page 53, 2006.
- Lindblom, B., Lubker, J., & Gay, T. Formant frequencies of some fixed mandible vowels and a model of speech motor programming by predictive simulation. *Journal of Phonetics*, 7:147-161, 1978.
- Zandipour, M. *A Modeling Investigation of Vowel to Vowel Movement Planning in Acoustic and Muscle Spaces*. Doctoral dissertation, Boston University, 2006.