Compensatory Response to Unexpected Jaw Perturbation Triggered by Formant Transitions During Speech

Mark K. Tiede[1,2*], Takayuki Ito[1*], David J. Ostry[3,1*]

[1]Haskins Laboratories
300 George Street, New Haven, CT 06511 USA

[2]Massachusetts Institute of Technology, Speech Communication Group – R.L.E.
50 Vassar Street, Cambridge, MA 02139 USA

[3]Department of Psychology, McGill University
1205 Dr. Penfield Avenue, Montréal, QC H3A 1B1 Canada

tiede@haskins.yale.edu, taka@haskins.yale.edu,
ostry@motion.psych.mcgill.ca

Abstract. *Observations were made of jaw and formant trajectories during speech perturbed by unexpected mechanical loads applied to the jaw. The perturbation forces were applied using a jaw-coupled robot, and triggered by a thresholding criterion applied to realtime tracking of the first formant. Subjects produced multiple repetitions of real word utterances containing either high-to-mid or low-to-mid vowel sequences. Perturbations were delivered one out of every five repetitions, selected at random, with half applied upward and half downward, and forces sustained throughout the target utterance. Audio and jaw position were recorded concurrently. Individual tokens were subsequently extracted using the F1 triggering threshold for alignment. Formants show initial deviation from control trajectories and then recovery that begins approximately 75 ms after the onset of perturbation. Compensation in most instances is nearly complete even though jaw position does not recover its unperturbed trajectory, and is thus presumably effected through appropriately modified tongue movements. This behavior is compatible with feedforward models of speech motor planning, in which corrective motor commands are computed in response to errors between anticipated and produced sensory consequences.*

[*]Supported by NIH/NIDCD

1. Introduction

A well-established means of probing the adaptability of the speech production system is to observe how it compensates under perturbation. In a pioneering bite-block study Lindblom *et al*. (1979) established that speakers succeeded in producing near-normal vowel formants despite the constraint on mandible position, and were able to do so immediately. Studies in which effects of unexpected perturbation of jaw position during bilabials were evaluated with respect to lip compensation (including Folkins & Abbs, 1975, Kelso *et al*., 1984, Gomi *et al*., 2002) have shown that the time course of such compensation is both rapid (e.g. 15-35 ms in OOI reported by Kelso *et al*.) and complete (the lips achieve closure). The Kelso *et al*. study also established through lip (OOI, OOS) and tongue (GG) EMG that compensatory response was selective and linked to the articulatory target: downward perturbations applied during the upward jaw cycle for the final /b/ in /baeb/ evoked lip but not tongue response, while perturbations of /z/ in /baez/ showed the opposite pattern.

However, it is difficult in EMG studies of this type to distinguish between the potential contributions of mechanical linkage between articulators, autogenic response, and true cortical feedback. By focusing on the perturbation of vowels, thereby limiting the kinds of cutaneous feedback associated with consonantal targets, we attempt here a preliminary investigation of compensation mediated principally by somatosensory and auditory feedback. Our approach is to apply unexpected loads to the jaw, in both upward and downward directions, that are triggered by an acoustic transition in the F1 trajectory. Since jaw position may vary as its stiffness increases during an experiment in response to perturbation, this acoustic triggering approach provides potentially greater consistency over methods linked to jaw height.

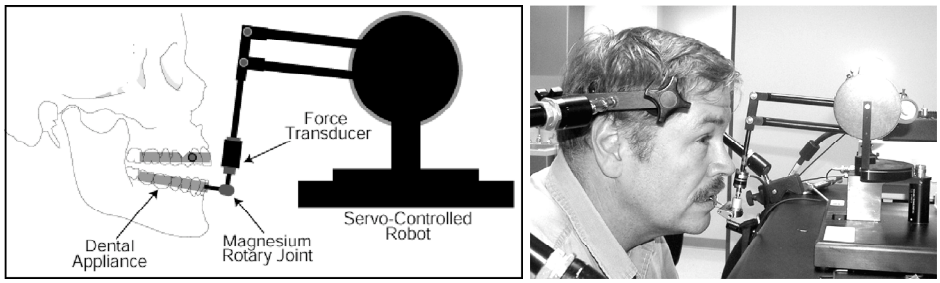

Figure 1. Experimental setup. The robot is coupled to the jaw using a subject-specific dental appliance; restraints minimize head movement.

2. Methods

2.1. Subjects

Two male native speakers of English (E1, E2) and one male native Japanese speaker (J1) participated. The two English speakers produced multiple repetitions of the utterances “see red” and “R.A.” for triggering respectively on low-to-mid F1 (jaw

opening) and high-to-mid F1 (closing) trajectories. The corresponding utterances produced by the Japanese subject were “ie” (house) and “Sae” (a name).

2.2. Experimental Setup

Mechanical forces were generated using a computer-controlled robotic device (Phantom 1.0, SensAble Technologies). Coupling to the jaw was made through an acrylic appliance (constructed separately for each subject from dental casts) which was glued to the buccal surface of the teeth, and attached to the robot by a rotary joint permitting unimpeded 6DOF movement. A force/torque sensor (ATI) intermediate between this joint and the tip of the robot was used to record resistive forces generated in response to perturbation. Head restraints mounted on adjustable quick-release support arms were used to constrain subject head movement. Audio was transduced using a directional microphone placed approximately 30 cm from the subject’s mouth, and digitized after hardware low-pass filtering using a 16 bit A/D device (PCI-6036E, National Instruments). A custom software program was developed to coordinate stimulus presentation, data recording, formant tracking, and control of the robot. The program also provided a realtime scrolling display of the audio signal and the trajectory of the first formant, to facilitate the choice of an appropriate triggering threshold. Speech sampled at 10 kHz was used to continuously update PARCOR coefficients of an adaptive lattice filter with exponential decay (iterative Burg algorithm; Orfanidis 1988), and formants were computed from associated predictor coefficients at 2 ms intervals by Levinson recursion.[1]

2.3. Experimental Procedure

Data were collected in blocks of 50 trials. Subject-generated (resistive) forces and the location of the jaw-coupled rotary connector joint relative to an initial calibration position were recorded during each trial at 1 ms intervals, together with concurrent audio recording at 10 kHz.

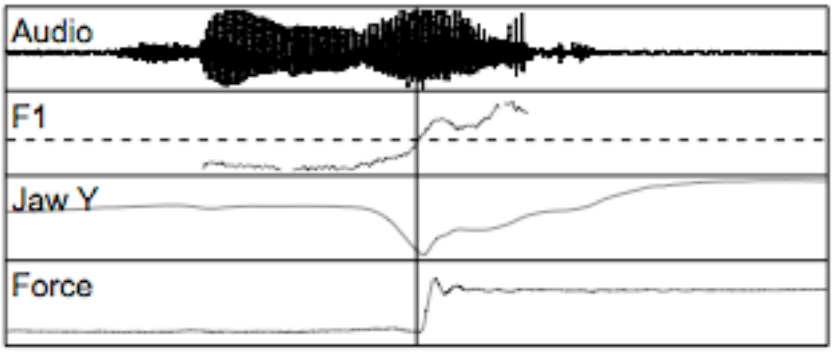

Figure 2. Example trial (“see red”) with an upward perturbation, showing F1 threshold, triggered force, and effect on jaw trajectory.

[1] This formant tracking approach follows the implementation of Reiner Wilhelms-Tricarico as used in the auditory perturbation experiment of Purcell & Munhall (2006).

At the onset of each trial the target utterance was displayed to the subject on a monitor as a cue to begin production.

Ten trials of each 50 were perturbed, half upwards and half downwards. In each block the perturbed trials were chosen randomly, subject to the constraint that no perturbed trial immediately follow another. During trials selected for perturbation, the robot was 'armed' by a sequence of F1 values that remained below (low-to-mid) or above (high-to-mid) the triggering threshold for a period of at least 50 ms, and the first F1 value exceeding this threshold initiated the perturbation. 3 Newton perturbation forces were generated along the vertical axis of the robot, applied sigmoidally with a 20 ms rise time, and sustained for one second (thus they were active for the remainder of each perturbed trial).

2.4. Data Analysis

Individual trials were extracted as 1000 ms intervals centered on the F1 triggering threshold for that utterance (whether used to trigger perturbation or not). In a few instances the tracking algorithm missed the correct F1 transition and incorrectly triggered the perturbation at a later, inappropriate point of the utterance; these cases were identified by interactive examination and removed from the dataset. Formant values were temporally aligned with jaw movement by subtracting the 18 ms latency of the iterative Burg filter (established with comparison to LPC analysis of the audio using sample-centered windowing, and consistent with the 20 ms value reported by Purcell & Munhall (2006) using a similar tracking algorithm). Formants were also processed with a cubic smoothing spline ($\rho = .5$) to ameliorate missing values and outliers.

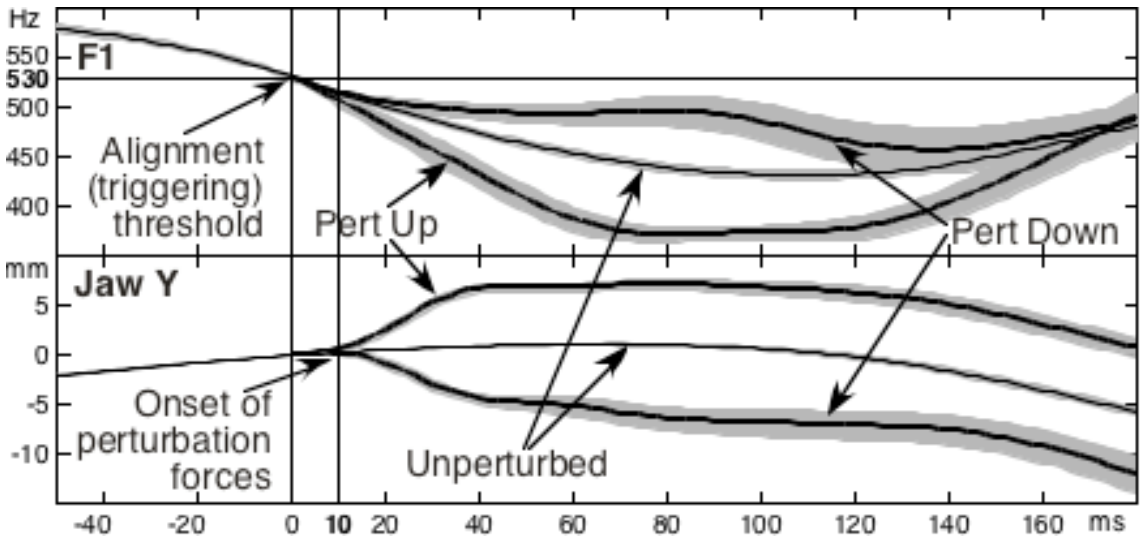

Figure 3. Aligned trials for subject E1 production of "R.A." showing mean trajectories +/- standard error at each sample. Note that F1 eventually recovers its unperturbed trajectory while the jaw does not.

Despite head bracing there was in some instances drift in the baseline jaw position, and so the vertical component of jaw movement for each trial was aligned spatially by subtracting its position at the F1 trigger alignment offset.

Measures of displacement for perturbed trials were obtained for both F1 and vertical jaw position on trajectories normalized by subtracting the mean of the corresponding unperturbed trajectories (see Figure 4). Subject data from different blocks with the same utterance type were analyzed together.

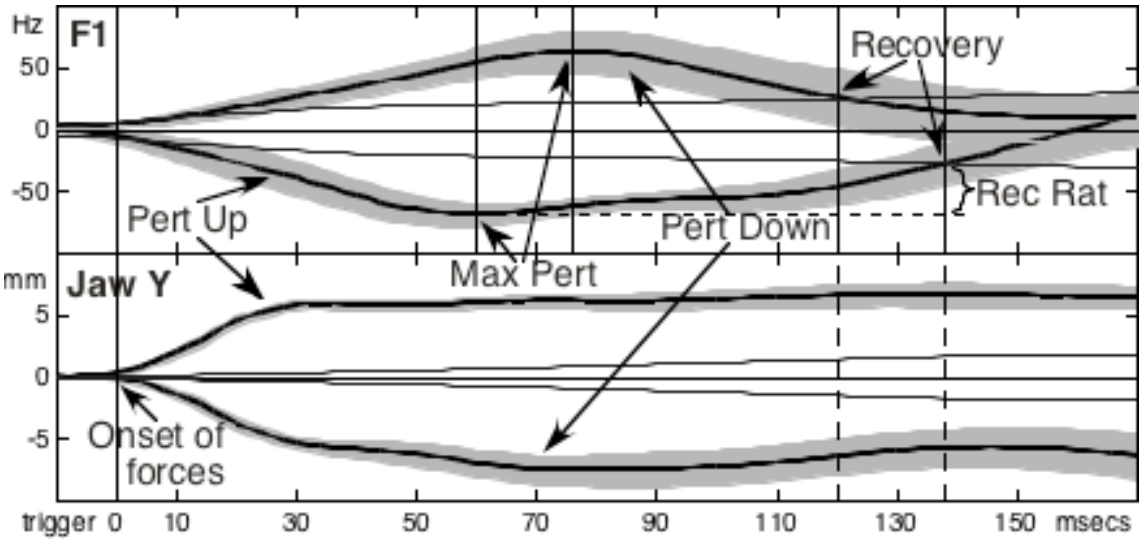

Figure 4. Measurements. Mean perturbed trajectories +/- std. error normalized by subtracting mean control trials, shown here for E1 utterance "R.A." Offsets measured relative to onset of forces at points of maximum F1 perturbation (**Max Pert**), and where mean trajectories recovered to within 1.5 std. dev. of control (**Recovery**). Ratio of values at Recovery : MaxPert (**Rec Rat**) shown for upwardly perturbed F1.

For each subject and utterance two temporal offset measurements were recorded relative to the onset of forces for each perturbation type: the point of maximum displacement in F1 of the mean perturbed from the mean control trajectory (*Max Pert*), and the point at which the mean perturbed trajectory recovered to within 1.5 standard deviation units of the control (*Recovery*). In addition, the ratio of the trajectory value at the recovery point to that of the F1 maximum displacement (*Rec Rat*) was obtained for both F1 and JawY.

3. Results

Within perturbed trials, the offset of the maximum displacement from the unperturbed control trajectory (*Max Pert*) represents the point at which formant correction was initiated. These timings, summarized in the first part of Table I, show that on average subjects began to correct for perturbation-induced deflections in F1 around 76 ms after the onset of the forces applied to the jaw. J1 (61) was quicker to react than either E1 (80) or E2 (88). Two subjects (E2, J1) reacted substantially more quickly for high-to-mid vowel utterances than low-to-high (25 and 13 ms; E1 differed by just 4 ms). Subjects responded to the direction of perturbation idiosyncratically: E2 slower to react on upward perturbations, J1 faster, E1 no difference.

The offset by which perturbed F1 trajectories returned to within 1.5 standard deviation units of the unperturbed control (*Recovery*) gives an indication of the time course of the compensatory response to perturbation. These timings are summarized in the second part of Table I. On average, subjects recovered to within the criterion 183 ms after the onset of forces, or 107 ms after the *Max Pert* point of maximum F1 displacement. Subject E1 (133) recovered in general more quickly than E2 (226) or J1 (191); however, all subjects compensated within about the same interval (149) for upwardly perturbed high-mid vowel stimuli.

A. Offsets of Maximum F1 Perturbation					
Vowel	Perturb	E1	E2	J1	Mean
high -> mid	up	90	74	54	73
	dn	74	76	54	68
low -> mid	up	70	110	58	79
	dn	86	90	76	84
	Mean	80	88	61	76

B. Offsets of F1 Recovery					
Vowel	Perturb	E1	E2	J1	Mean
high -> mid	up	152	142	152	149
	dn	102	340	242	228
low -> mid	up	148	220	172	180
	dn	130	200	196	175
	Mean	133	226	191	183

Table I. Offsets of maximum F1 displacement of averaged perturbed trials from unperturbed baseline (A), and time to recovery within 1.5 s.d. of baseline (B); all timings in ms relative to onset of forces.

The perturbed trajectories were sampled at both the *Max Pert* and *Recovery* offset points to form a ratio (*Rec Rat*) for comparing the extent of compensation in both F1 and vertical jaw position. Figure 5 shows the values at recovery as percentages of the maximum displacement values. F1 values at *Recovery* are on average 43% of their maximally perturbed values as compared to an average 96% for Jaw Y, indicating that the recovery in F1 occurs for the most part independently of jaw position.

4. Discussion

The principle focus of this work has been on development of the acoustic triggering method for jaw perturbation; accordingly the amount of collected data is small, and intended primarily for validation of the technique. Nonetheless, the compensation results obtained are consistent with earlier studies of unexpected perturbation. For example, the range of reaction times (*Max Pert*) representing the beginning of compensation (61 – 88 ms) is comparable to the 72 – 164 ms range reported by Honda *et al*. (2002) induced using their inflatable artificial palate.

The timing differences observed in *Max Pert* between subjects E1 (80) and E2 (88) *vs*. J1 (61) are potentially due to the /r/ included in the English stimuli, intended (unnecessarily, as the Japanese data show) to accentuate the formant transition used for triggering: it may be that perturbations applied during the transitional rhotic take additional time to be sensed as problematic.

One especially clear result, illustrated by Figures 4 and 5, is that the mechanism for effecting compensation in F1 does not involve the jaw. Despite the relatively low force applied (3 N) the jaw does not recover its unperturbed trajectory, and so formant recovery is presumably effected through adjustments in tongue posture.

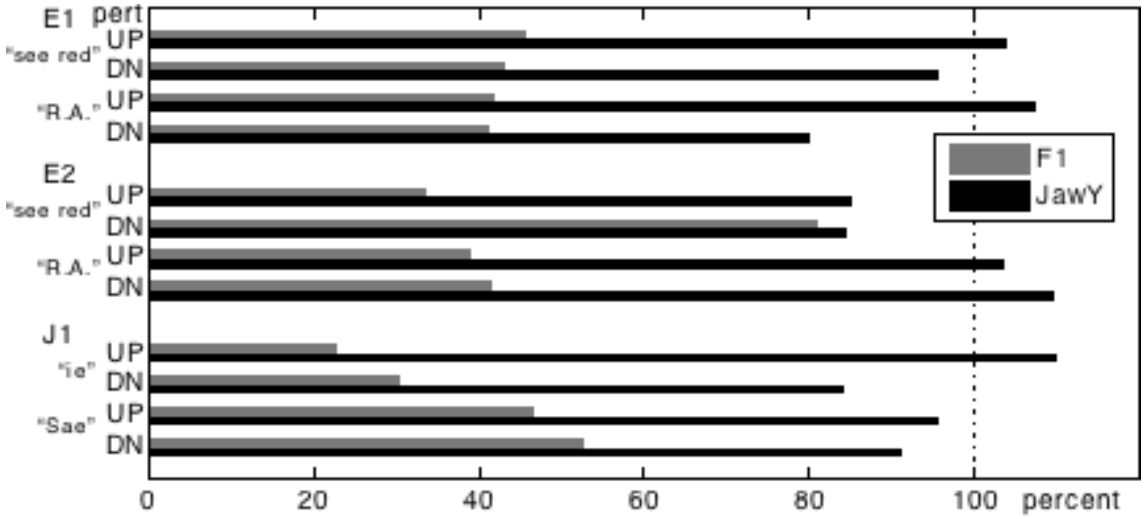

Figure 5. Ratio of trajectory values at point of recovery to point of maximum perturbation (Rec Rat). At Recovery F1 values have dropped to an average 43% of their maximally perturbed values, while JawY values averaging 96% remain close to their perturbed maxima.

The time course of compensation observed here is consistent with goal-directed models of speech motor planning that include a feed-forward component (Lindblom *et al*. 1979, Kelso *et al*. 1984, Saltzman 1986, Guenther 1995). In these models, a "flexibly assembled coordinative structure" (Kelso et al. 1984, p. 812) is associated with predicted sensory input. When a mismatch is detected between anticipated and produced sensory consequences, corrective motor commands are computed as necessary to achieve the goal. Because in this study applied forces persist throughout the duration of perturbed stimuli, the effects of such correction can be observed.

5. Summary

This work represents a preliminary effort to establish a new technique for applying unexpected loads to the jaw during speech, triggered through transitions observed in real-time monitoring of the first formant. Perturbation forces applied in two directions were successfully and consistently triggered, and made to both jaw opening (high-to-mid vowel) and jaw closing (low-to-high vowel) targets. Results show initial deviation from control trajectories, and then recovery in F1 (but not jaw height) that begins about

76 ms after the onset of perturbation. Compensation in F1 is in most cases nearly complete after an additional 107 ms. Since the jaw does not recover its unperturbed trajectory this is presumably effected through appropriately modified tongue movements. The observed compensatory behavior is compatible with feedforward models of speech motor planning, in which corrective motor commands are computed in response to errors between anticipated and produced sensory consequences.

References

Folkins, J. W. and Abbs, J. H. Lip and jaw motor control during speech: responses to resistive loading of the jaw. *Journal of Speech and Hearing Research*, 18(1): 207-220, 1975.

Gomi, H., Honda, M., Ito, T., and Murano, E. Z. Compensatory articulation during bilabial fricative production by regulating muscle stiffness. *Journal of Phonetics*, 30: 261-279, 2002.

Guenther, F. H. Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, 102: 594-621, 1995.

Honda, M., Fujino, A., and Kaburagi, T. Compensatory responses of articulators to unexpected perturbation of the palate shape. *Journal of Phonetics*, 30: 281-302, 2002.

Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E., and Fowler, C. A. Functionally specific articulatory adaptation to jaw perturbations during speech: evidence for coordinative structures. *Journal of Experimental Psychology*, 10(6): 812-832, 1984.

Lindblom, B., Lubker, J., and 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, 1979.

Orfanidis, S. J. Optimum Signal Processing: An Introduction. McGraw-Hill, 2nd Edition, 1988.

Purcell, D. W. and Munhall, K. Compensation following real-time manipulation of formants in isolated vowels. *Journal of the Acoustical Society of America*, 119(4): 2288-2297, 2006.

Saltzman, E. Task dynamic coordination of the speech articulators: a preliminary model. *Experimental Brain Research*, 15: 129-144, 1986