Speech Communication – Speech Motor Control

Visiting Scientists and Research Affiliates

Dr. Jana Brunner, Humboldt University, Berlin, Germany
Dr. Deryk Beal, Boston University

Constraints and Strategies in Speech Production

Introduction

The objective of this research has been to refine and test a theoretical framework in which words in the lexicon are represented as sequences of segments and syllables and these units are represented as complexes of auditory/acoustic and somatosensory goals. The motor programming to produce sequences of sensory goals utilizes an internal neural model of relations between articulatory motor commands and their acoustic and somatosensory consequences. The relations between articulatory motor commands and the movements they generate are influenced by biomechanical constraints, which include characteristics of individual speakers' anatomy and more general dynamical properties of the production mechanism. To produce an intelligible sound sequence while accounting for biomechanical constraints, speech movements are planned so that sufficient perceptual contrast is achieved with minimal effort. There are individual differences in planning movements toward sensory goals that may be due to relations between production and perception mechanisms in individual speakers.

In the later years of the project, the internal model was implemented as a neurocomputational model that was used to control a vocal-tract model (an articulatory synthesizer). The combined models provide the bases of hypotheses about the planning of speech movements. To test these hypotheses, we conducted experiments with speakers and listeners in which we measured articulatory movements, speech acoustics, speech perception, and brain activation. We manipulated speaking condition, phonemic context and speech sound category and we introduced transient and sustained perturbations.

After a year-long unfunded extension the project ended officially 11/30/09. Since then, we have completed studies that were underway and have continued to report results at meetings and in publications.

1. Auditory feedback and the control of the spatial aspect of articulation

Articulation of a multisyllabic utterance consists of two aspects: a) the spatial aspect, which involves the spatial positions of the articulators during different phonemes of the utterance, and b) the temporal aspects, which refer to the times at which those successive phonemes are realized. In these two experiments, we separately examined the role of AF in these two aspects of articulatory control. This was achieved by using the all-vocalic sentence “I owe you a yo-yo” as the stimulus utterance. Due to a continuous glottal excitation (voicing) during this sentence, formant trajectories can be tracked relatively accurately during the entirety of this sentence (See Fig. 1.1), which brings two advantages. First, since formant frequencies reflect positions of the articulators, they can be used as surrogates for the articulator positions, allowing us to bypass the more involved procedure of articulographic recordings. Second, it allows us to extract the timing of the syllables from the acoustic recording alone. In the following, we describe in detail the methods we used to extract articulatory position and timing from the second formant (F2) trajectory during this utterance.

In Study 1, we aimed to test whether AF is involved in the online control of the spatial aspect of multisyllabic articulation by measuring changes in formant trajectories caused by perturbations to speakers' feedback of F2 during production of the stimulus utterance.

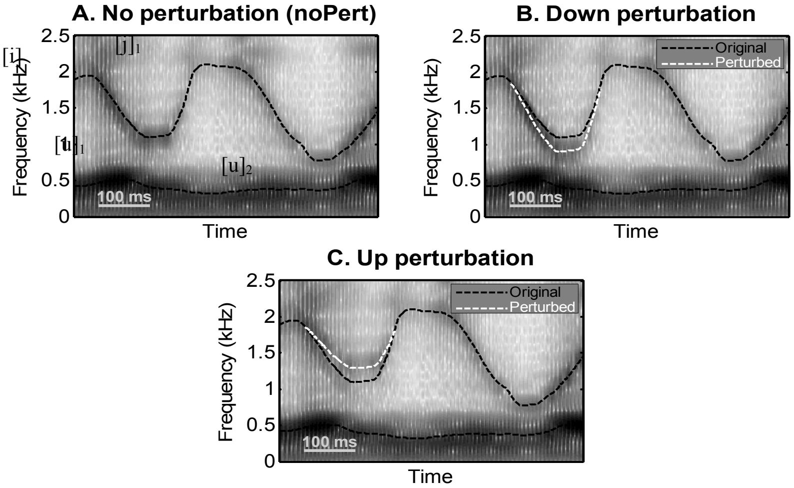


Figure 1.1. Examples of the spectral perturbations used in Study 1. A: original spectrogram of a production of the sentence “I owe you a yo-yo”; only the part of the utterance during the words “owe” and “you” is shown. **B**: Downward (*Down*) perturbation auditory feedback of the sound in A. **C**: Upward (*Up*) perturbation. *Dashed black traces*: original F1 and F2; *dashed white traces*: perturbed F2.

Nineteen healthy adult native English speakers participated in this study. The subjects were trained to produce the stimulus utterance within durations between 1.2 and 1.6 seconds, which was within the range of natural speaking rates.

A custom computer program was used to perturb the trajectory of F2 between the [i] sound in “I” and the [j] sound in “you” and. The F2 perturbation was based on time-varying digital filtering. Two types of perturbation were delivered: (1) *Down* perturbation, in which the extent of the F2 transition was increased by approximately 25% relative to the subject’s production (see Fig. 1.1B for an example), and (2) *Up* perturbation, in which the F2 transition was decreased by approximately 25% (Fig. 1.1C). These perturbations were called spectral perturbations because they alter the magnitudes of the formant frequencies at the F2 minimum without changing the timing of the minimum. The smoothness of the formant changes was ensured in both types of perturbations, in order to avoid noticeable formant discontinuities. An experiment consisted of 120 repetitions of the utterance, arranged into 20 blocks of six. In each block, there was one *Down* trial and one *Up* trial; the rest were unperturbed (*noPert*) trials. Order of the three types of trials was randomized.

As Fig. 1.2A shows, the grand averages of the normalized F2 trajectories across the productions of the 19 subjects show divergent patterns under the three different conditions (*noPert, Down* and *Up*).

Spatial effects: Subjects made formant frequency changes during and after the perturbed syllable in response to the perturbations. We found that the time-varying spectral perturbations of the F2 transitions during the multisyllabic utterance led to spatial (spectral) compensations in the subjects’ articulation. To isolate the spatial effects of the perturbations from the temporal effects, we computed the changes in the absolute value of F2 at a few landmarks in the F2 trajectory. These landmarks included the F2 minimum at [u]1, the temporal midpoint between the F2 minimum in [u]1 and the F2 maximum in [j], the maximum in [j] (dubbed the [u]1-[j]1 midpoint), and the temporal midpoint between the maximum in [j] and the following minimum in [u]2 (dubbed the [j]1-[u]2 midpoint).

No significant effect of the perturbations was observed at the F2 minimum in [u]1 (Fig. 1.2D). This result is not surprising given that this F2 minimum is relatively close in time to the onset of the perturbation. A significant difference between the *Down* and *Up* conditions was observed at the

[u]1-[j]1 midpoint (p<0.05, repeated-measures ANOVA followed by post-hoc Tukey test). The Down-perturbed trials show significantly higher F2 value compared to the Up-perturbed trials (Fig. 1.2E). The F2 maximum in [j] showed a similar trend, which also achieved statistical significance (Fig. 1.2F). At the [j]-[u]2 midpoint, which is approximately 100 ms after the end of the perturbation, a significant effect of the perturbation could still be observed (Fig. 1.2G).

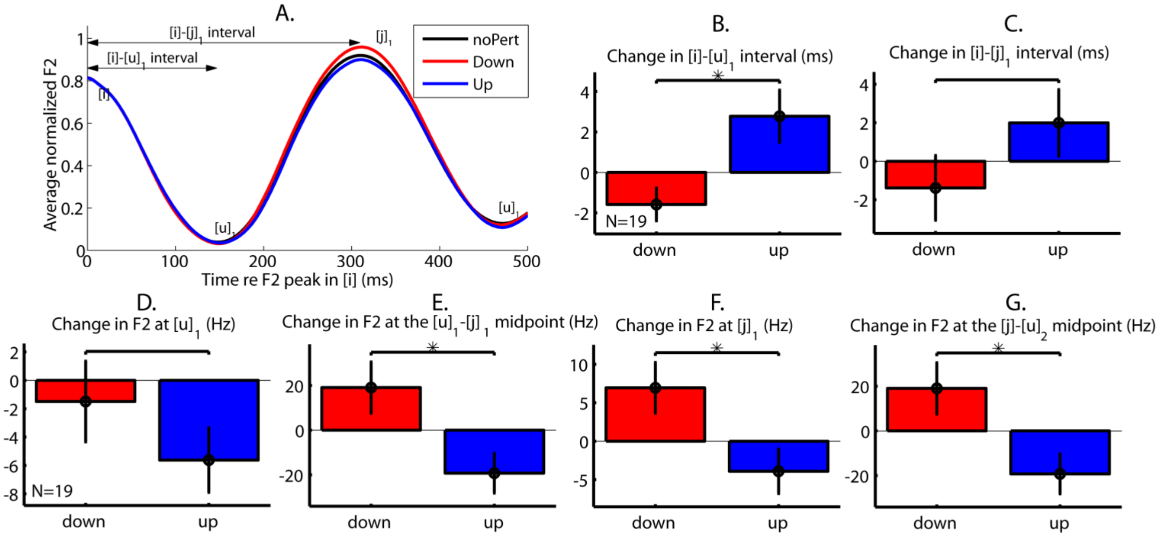


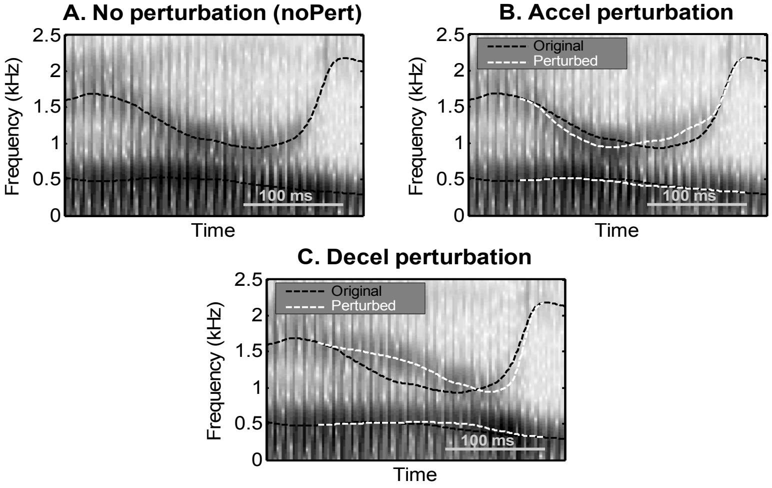
Figure 1.2. Articulatory compensations under the spectral (*Down* and *Up*) perturbations. A: grand average (across trials and subjects) of time-normalized F2 trajectories aligned at the F2 maximum in [i] of "I". Error bands are omitted for clarity of visualization. F2 was normalized in order to permit averaging across subjects. **B** and **C:** timing changes under the perturbations. **B**: change in the time interval between the F2 maximum in [i] and the F2 minimum in [u] of "owe" ([u]$_1$) (mean±1 standard error). **C**: change in the interval between the F2 minimum in [u]$_1$ and the F2 maximum in [j] of "you. **D – G**: spatial (spectral) changes under the perturbations. Note that the y-scales are different between these plots. **D**: change in the value of F2 at the minimum in [u]$_1$. **E**: change in value of F2 at the temporal midpoint between the F2 minimum in [u]$_1$ and the F2 maximum in [j]. **F**: change at the F2 maximum in [j]. **G**: change at the midpoint between the F2 maximum in [j] and the F2 minimum in [u] of "you" ([u]$_2$). Asterisks: significant difference between the *Down* and *Up* conditions (p<0.05, post-hoc Tukey test following repeated-measures ANOVA).

Temporal effects: The spatial perturbations had influences on timing, which were greater for intra-syllabic timing than for inter-syllabic timing. To study the temporal effects of the perturbations, we examined the time interval between the F2 maximum during [i] (in the word "I") and the F2 minimum during [u]1 ([u] in "you"), dubbed the [i]-[u]1 interval. The [i]-[u]1 interval is roughly the duration of the syllable "owe", and is hence a measure of intra-syllabic timing. The average changes in the [i]-[u]1 interval from the *noPert* baseline are shown in Fig. 1.2B. The *Down* and *Up* trials showed decreased and increased [i]-[u]1 intervals, indicating earlier and later termination of the syllable "owe" relative to the control (*noPert*) condition, respectively. The difference between *Down* and *Up* was statistically significant (p<0.05, post-hoc Tukey test). The interval between the F2 maxima corresponding to [i] in "I" and [j] in "you", called the [i]-[j] interval, is a measure of inter-syllabic timing. It showed trends similar to the [i]-[u]1 interval. However, unlike the [i]-[u]1 interval, this change in the [i]-[j] interval was not statistically significant. (Fig. 1.2C).

Thus, it can be seen that the spatial compensations were manifested as changes in F2 in the subjects' production in directions opposite to the auditory perturbations. These spatial compensations were found not only during the perturbation, but also in the production after the perturbation, which was indicative of a predictive compensation that utilizes AF from earlier segments to guide the control of subsequent segments. As for temporal effects, timing compensations observed under the *Up* and *Down* perturbations in this experiment were mostly local, in the sense that they were confined to the timing of intra-syllabic events and had no significant effects at the inter-syllabic level.

2. Auditory feedback and the control of the temporal aspect of articulation

Study 1 demonstrated a role of AF in spatial control of articulation. However, it left unanswered the question of whether AF is involved in controlling inter-syllable timing. Significant effects of the perturbations on inter-syllabic timing were not observed in Study 1. It is possible that the speech motor system does not rely on AF for controlling inter-syllabic timing. Alternatively, it is possible that the *Down* and *Up* perturbations were not appropriate for eliciting inter-syllabic timing changes, perhaps because those perturbations did not alter the timing of the F2 extrema. In order to distinguish between these two possibilities, we conducted a second study (Study 2). In this study, we deployed two new types of perturbations of the F2 trajectory, which are called temporal perturbations. These differed from the spectral perturbations used in Study 1, in that they altered the timing of the F2 minimum in the [u]1 sound (in word "owe") without significantly altering the magnitude of F2. Examples of the two types of temporal perturbations: Acceleration (Accel) and Deceleration (Decel) are shown in Fig. 2.1.


A. No perturbation (noPert)

B. Accel perturbation

C. Decel perturbation

Figure 2.1. Examples of the temporal types of AF perturbations. These are examples from a recording of the perturbations. The part of the utterance illustrated in the spectrograms starts from the end of the word "I", and ends at the middle of the word "owe". The perturbations are activated between the [i] sound in the word "I" and the [j] in the word "you". **A**: spectrogram of the un-perturbed (original) signal. **B**: spectrogram of the acceleration (*Accel*) perturbation; **C**: deceleration (*Decel*) perturbation. The *Accel* and *Decel* perturbations manipulate the temporal aspects of AF, whereas the *down* and *Up* perturbations alter the spectral aspects of AF.

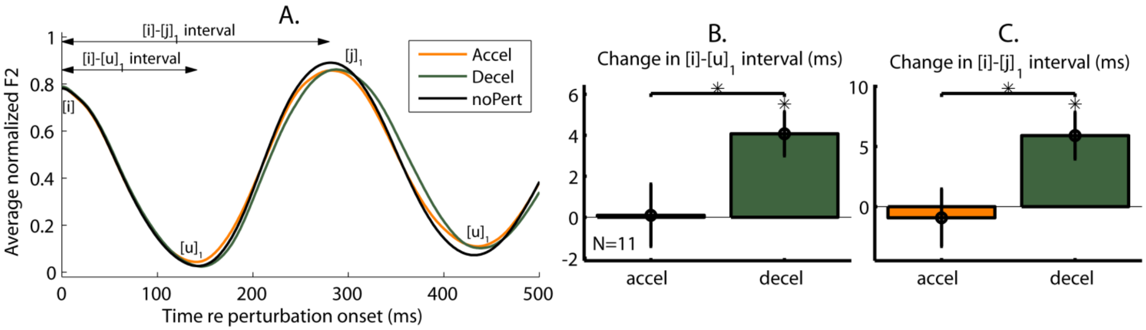


Figure 2.2. Articulatory compensations under the temporal (*Accel* and *Decel*) perturbations. A: grand average (across trials and subjects) of time-normalized F2 trajectories aligned at the F2 maximum in [i] of "I". The format is as in Fig. 1.2A. **B** and **C:** timing changes under the perturbations. **B**: change in the time interval between the F2 maximum in [i] and the F2 minimum in [u] of "owe" ([u]$_1$) (mean±1 standard error). **C**: change in the interval between the F2 minimum in [u]$_1$ and the F2 maximum in [j] of "you. Asterisks: significant difference between the Down and *Up* conditions ($p<0.05$, post-hoc Tukey test following repeated-measures ANOVA).

A new group of 11 subjects participated in this study. The experimental design was identical to that in Study 1, except that the temporal perturbations were used instead of the spectral ones. The grand averages of F2 trajectories produced by the subjects under the baseline (*noPert*) and the two perturbation conditions are shown in Fig. 2.2A. It can be seen in these trajectories that the timing of the F2 minimum in [u]1 and that of the F2 maximum in [j]1 were both delayed under the *Decel* perturbation compared to the baseline production, whereas the timing of these extrema under the *Accel* perturbation do not appear to differ substantially from the baseline. Quantitative analyses of the changes in the [i]-[u]1 interval and the changes in the [i]-[j]1 interval are shown in Fig. 2.2B and C. These analyses confirmed the asymmetric effects of the *Accel* and *Decel* perturbations: while no significant timing changes occurred under the *Accel* perturbation, *Decel* perturbations led to significant lengthening of both time intervals relative to the baseline production. Despite the relatively small magnitude of these compensatory timing shifts (~4-6 ms), these observations do indicate a role of AF in the online control of syllable timing at both the intra- and inter-syllabic levels, thus addressing the question left from Study 1.

2. Effects of Hearing Status on Adult Speech Production

This project aims to advance knowledge of the roles of hearing in speech and to evaluate new hypotheses based on a model of those roles. We are using acoustic recordings of speech, perceptual tests, and several types of intervention in experiments with individuals with normal hearing and individuals who have postlingual deafness and receive cochlear implants. According to our model, the goals of speech movements are in sensory domains. Motor commands to achieve auditory goals are determined by the combined operation of feedback and feedforward control mechanisms. Feedforward control is almost entirely responsible for generating articulatory movements in mature speech production. However, if auditory feedback detects a mismatch between auditory goals and the heard consequences of ongoing speech movements, the detected error leads to the generation of feedback-based corrective motor commands, which in turn serve to update feedforward commands for subsequent movements.

In our experiments, the usefulness of a speaker's auditory feedback is reflected in that speaker's acuity for the speech parameters under study, based on studies indicating that the size and spacing of the speaker's phonemic goal regions can be indexed by measures of the speaker's produced phonemic contrasts. The relative contributions of feedback and feedforward control are assessed by blocking and unblocking feedback, and by introducing interventions and measuring speakers' compensatory responses. The main objective of the proposed research, then, is to extend our understanding of relations among speakers' auditory acuity, the phonemic contrasts they produce, and the roles of feedback and feedforward control. To reach this objective we are engaged in a series of experiments involving: 1) *Auditory acuity and produced phonemic contrast,* 2) *Auditory acuity and produced lexical stress,* 3) *Congruence of vowel spaces in production and perception* and 4) *Vowel imitation and auditory feedback*. The resulting comprehensive picture should provide significant new insights into the roles of speaker acuity in feedback and feedforward control of speech motor planning to achieve auditory goals.

We are currently completing a full set of studies on groups of 21 cochlear implant users and 21 age- and gender-matched normal-hearing controls. A substantial proportion of the data has been gathered, processed and analyzed – enough to provide a considerably more advanced report of the results. These results are described below in terms of hypotheses we are testing and the outcomes to date. Background, methods and earlier results have been reported previously.

2.1. Phoneme goals and auditory acuity.

Hypothesis: Hearing speakers, compared to implant users, will show greater produced phoneme contrasts and smaller production goal regions.

Result: Hearing speakers had greater contrasts than implant users for one of the two vowel contrasts tested, as hypothesized, but not the other. Also as expected, hearing speakers had smaller goal regions than implant users for both contrasts.

Hypothesis: Acuity in vowel discrimination will be greater for hearing speakers than implant users. Speakers with greater acuity will show a greater degree of phoneme contrasts and smaller vowel production goal regions.

Result: As hypothesized, for both vowel contrasts, hearing speakers had greater acuity in discriminating changes in vowel quality than did implant users. However, in contrast to prior findings, we found no evidence in this study that speakers with greater acuity in vowel discrimination produced more marked vowel contrasts. Speaker acuity did correlate with the size of the speaker's vowel production goal regions for one vowel contrast but not the other. In addition, speaker acuity correlated with overall accuracy of vowel identification in one of the two contrasts.

2.2. Production of lexical stress.

Hypothesis. Compared to speakers with normal hearing, implant users will show less marked parameter contrasts in lexical stress production.

Result: For both groups, normalized duration (stressed / unstressed syllable) is the parameter that showed the greatest contrast between nouns and verbs in lexical stress. Whereas the longest normalized durations occur on the verbs (second syllable stressed), the greatest pitch changes occur on the nouns. Normalized SPL changes were greater for the verbs. As hypothesized, implant users followed the same pattern of results obtained from hearing speakers, but they make more modest parameter changes when producing lexical stress.

2.3 Congruence of vowel spaces in production and perception.

Hypothesis: There will be significant correlations among perceptual and production goal regions for vowels.

Result: With six vowels, the 15 inter-vowel distances in perception were correlated with those in production for each speaker and group averages of correlation coefficients were computed. Hearing speakers and implant users had about the same high degree of agreement between inter-vowel distances in production and perception. When speakers' intervowel distances in production were predicted from their intervowel distance in perception and vowel pair, the identity of the vowel pair made a significant contribution to the multiple correlation, for both hearing speakers and implant users. This result indicates that the speaker produces a vowel space depending on the perceived vowel space, consistent with the view that perception and production have the same vowel targets.

2.4. Auditory Feedback and Vowel Imitation.

Hypothesis: When the speaker imitates synthesized vowels, the formants of his or her matching vowels can be predicted from the target to imitate and from the speaker's productions of those vowels without a target ("canonical vowel formants"). Canonical formants will have more weight when imitation is conducted without auditory feedback.

Result: Imitation formants were predicted reliably. Canonical vowels received substantial weight but not more weight without feedback.

Hypothesis: Hearing speakers' vowel imitations will be closer to their targets than those of implant users. Vowel imitations will be significantly further from their targets when conducted without auditory feedback.

Result: As hypothesized, hearing speakers' vowel imitations were closer to their targets than those of implant users and both groups' accuracy of imitation was facilitated by auditory feedback. We also observed that the variability of imitations around the target was greater for the implant users and greater without feedback than with it.

Hypothesis: The greater a speaker's ability to discriminate subtle changes in vowel spectra (acuity), the closer his or her imitations will be to the target. The correlation will be higher with auditory feedback than without.
Result: As hypothesized, the acuity of a speaker was significantly correlated with that speaker's proximity to the target in vowel imitations, and the correlation was greater with feedback than without.

Hypothesis: With auditory feedback unavailable, imitation responses will be nearer canonical values.

Result: This hypothesis is unsupported. When feedback was removed and imitations moved away from targets, those imitations were also further from canonical values, contrary to the hypothesis. Similarly, the data contradicted our expectation that imitation responses without feedback would cluster more around canonical values than those with feedback.

Hypothesis: High acuity speakers will have less dispersion of formant values around their mean imitation responses compared to low acuity speakers.

Result: As hypothesized, pooling over 19 speakers, two feedback conditions and two synthesized vowel continua, the speaker's acuity was significantly correlated with the accuracy of that speaker's imitations (distance from target).

3. Neuroanatomical and behavioral anomalies in persistent developmental stuttering

Persistent developmental stuttering is a disorder that affects 1% of the population and can have serious social and psychological consequences. This aim of this project was to demonstrate relations between measures of structural anomalies that have been identified in the brains of persons who stutter and measures of their speech movements and acoustics, and to make comparisons with the same measures from people with fluent speech. We applied for an R01 grant last year for this work, but did not qualify for funding. As we were preparing our only possible resubmission at the beginning of 2010, we received $100,000 in direct costs in the form of an R56 award from NIDCD, because stuttering is congressionally mandated high-priority research area and because of our group's research record. In spite of this expression of support from NIDCD, the resubmission was not reviewed well enough to qualify for funding.

In the spirit of the R56 award, we are beginning to run an initial study that was described in the R01 application. If successful, this study should validate the proposed methods and approach and make it possible for future researchers to pursue this line of investigation. Accordingly, we have used part of the funds to buy a new movement transducer system that will make it possible for us to measure articulatory movements from persons who stutter in less time and with less intrusiveness than with our previous system. We plan to use the remaining funds from the R56 award to support the other costs involved in performing the study.

Publications

Journal Articles, Accepted for Publication

Brunner, J., Ghosh, S.S., Hoole, P., Matthies, M., Tiede, M. & Perkell, J. (in press). The influence of auditory acuity on acoustic variability and the use of motor equivalence during adaptation to a perturbation. *Journal of Speech, Language, and Hearing Research*.

Cai S, Ghosh SS, Guenther FH, and Perkell JS. (In press). Adaptive auditory feedback control of the production of the formant trajectories in the Mandarin triphthong /iau/ and its patterns of generalization. *Journal of the Acoustical Society of America*.

Ghosh, S.S., Matthies, M.L. Maas, E. Hanson, A. Tiede, M. Ménard, L. Guenther, F.H.. H Lane and Perkell J.S. (in press). An investigation of the relation between sibilant production and somatosensory and auditory acuity, *Journal of the Acoustical Society of America*.

Perkell, J.S. Movement goals and feedback and feedforward control mechanisms in speech production, *Journal of Neurolinguistics* (invited submission for a special issue, "Neural Theory of Language")

Meeting Papers, Presented

Brunner, J, Ghosh, SS, Hoole, P, Matthies, ML, Tiede M, and Perkell, J. Relationship between auditory acuity and the use of motor equivalent strategies. Motor Speech Conference, Savannah, Georgia, March, 4-7, 2010.

Brunner, J, Hoole, P, Guenther, FH, and Perkell JS. Dependency of compensatory strategies on the shape of the vocal tract during speech perturbed with an artificial palate. *Journal of the Acoustical Society of America,* 127(3), 1854A. 159th Meeting of the Acoustical Society of America, Baltimore, MD, April 19-23, 2010.

Cai S, Ghosh SS, Perkell JS, and Guenther FH. The role of auditory feedback in the online control of articulatory trajectories and timing in a multi-syllabic utterance. Motor Speech Conference, Savannah, GA, March 4 - 7, 2010

Cai S, Ghosh SS, Guenther FH, and Perkell JS. Coordination of the first and second formants of the Mandarin triphthong /iau/ revealed by adaptation to auditory perturbations. (Abstract) J. Acoust. Soc. Am. 127(3), 2018. The 159th Meeting of the Acoustical Society of America, Baltimore, MD, April 19 - 23, 2010.

Matthies, ML, Perkell, JS, & Tiede, M (2009) Sources of feedback in the /s/-/sh/ contrast. Annual Convention of the American, Speech, Language and Hearing Association, New Orleans LA, session 2513, pg 153A.

Perkell, JS, Hisagi, M, Lane, H, Tiede, MK, Matthies, ML, Hanson, A, Ghosh, SS, & Guenther FH Auditory feedback and vowel imitation, Motor Speech Conference, Savannah, Georgia, March, 4-7, 2010.

Meeting Papers, Published

Brunner, J., Hoole, P., Guenther, FH. & Perkell, JS. (2010). Dependency of compensatory strategies on the shape of the vocal tract during speech perturbed with an artificial palate. *Proceedings of Meetings on Acoustics* 9.