Partial compensation in speech adaptation

Shira Katseff and John Houde

Abstract

We propose that speech planning is driven by targets which incorporate both auditory and somatosensory feedback. In three experiments, we show that (1) talkers compensate for abnormal auditory feedback, (2) compensation is more complete for small shifts in feedback than for large shifts in feedback, and (3) the completeness of compensation is driven by the discrepancy between auditory and somatosensory feedback.

Introduction

Human speech is dynamically adaptable. We can change our speech style for new audiences, restart words and phrases to repair errors, and learn to produce new sounds. None of these activities would be possible without a sophisticated system for tracking and correcting running speech. The question of what it is that we are tracking and correcting is essential to understanding the units of the speech we produce.

During speech, the central nervous system receives a large amount of realtime information about the acoustics and articulation of the words we produce. Talkers receive feedback from mechanoreceptors on the tongue, the positions of articulators in space, equilibrium points for muscles, and pitch receptors on the vocal fold mucosa (Wyke, 1983; Sanguineti, Laboissière, & Ostry, 1998; Pascal Perrier & Payan, 1996; Guenther & Barreca, 1997; Shiba, Miura, Yuza, Sakamoto, & Nakajima, 1999). Signals from the auditory cortex are integrated with this somatosensory information at an early stage of processing (Schroeder et al., 2001). Thus, a speech control system has access to a variety of tuning parameters in its decisions about adjustments in running speech. For the purposes of this paper, we split the many sources of feedback into two types: somatosensory and acoustic (also called 'perceptual'). Current models of speech control, for example DIVA (Guenther, 1995) incorporate both of these but do not address their relative importance.

Responses to abnormal feedback

One way to find out how much of this information is monitored by the auditory system on an ongoing basis is to alter it. Experimental evidence suggests that talkers are sensitive to perturbations in both auditory and somatosensory feedback. In experiments that alter auditory feedback, talkers repeat a target word or single sound and hear their voice resynthesized through headphones as they speak. Speech resynthesis occurs in real time and initially this feedback is an accurate reflection of the talker's voice. Over the course of the experiment, a critical attribute of the auditory feedback is altered.

Talkers compensate by opposing these feedback alterations. When vowel formants in auditory feedback are raised, talkers lower those formants in their speech. Likewise, talkers raise their vowel formants when those formants are lowered in their auditory feedback (Houde & Jordan, 2002; Purcell & Munhall, 2006). This general result has been replicated for F0 (Burnett, Freedland, Larson, & Hain, 1998) and for non-English speakers (Jones & Munhall, 2005).

Past experiments have altered auditory feedback in a strictly increasing fashion up to some maximum "plateau" value. From such experiments we have an idea of how much talkers will compensate for 200 Hz shifts in F1 and F2 and to 1 semitone shifts in F0. Notably, none of the experiments ever show more than 50% average compensation. That is, if an average subject's F1 feedback is raised by 200 Hz, that subject will lower F1 in the vowels s/he produces by about 100 Hz. Compensation during speech adaptation is incomplete and variable.

Summary of recent auditory speech adaptation experiments

Research Group	Formant	Amount of Shift	Language	% Compensation
Jones & Munhall, 2005	F0	1 semitone	Mandarin	40%
Jones & Munhall, 2000	F0	1 semitone	English	25%
Purcell & Munhall, 2006	F1	200 Hz	English	25%
Pile, Dajani, Purcell, & Munhall, 2007	F1 or F2	200 Hz	English	22.5%
Houde & Jordan, 2002	F1 & F2	200 Hz	English	50%

Compensation of any sort is evidence that subjects are monitoring their auditory feedback. However, that subjects do not compensate fully is perhaps puzzling: if talkers are sensitive to their auditory feedback and perfectly able to change their vowel formants to oppose the feedback shift, why should they stop at 25% compensation? It appears that reaching their targets does not require perfect alignment with their baseline vowel formants. There are at least these plausible explanations for this observation:

1. The target is fuzzy.

   Perhaps the range of acceptable vowel productions is large. This hypothesis predicts that compensation for similar shifts in auditory feedback should be indistinguishable and that the direction of compensation should not be predictable.

2. Bone conduction is a critical part of the target.

   It is possible that subjects are hearing their true vowels through bone conduction, attenuating the response to auditory feedback. This hypothesis predicts that silent and whispered speech should exhibit complete compensation.

3. The target is auditory, but is only indirectly related to the formant being shifted.

   Compensation might be complete if it were considered as a function of multiple formants.

4. The target is partly somatosensory.

   Perhaps targets have both acoustic and somatosensory components. A partly somatosensory target has some support from previous literature (Larson, Altman, Liu, & Hain, 2008). In a representative experiment, Tremblay, Shiller, & Ostry (2003) pulled the jaw forward while talkers were producing speech and nonspeech. They find that, while compensation is eventually complete for speech sounds, nonspeech sounds do not show complete adjustment. Complete compensation is even present in silently produced speech.

The following three experiments sought to distinguish between these possibilities.

Experiment 0: Replication

Methods

The first experiment sought to replicate the results of Purcell & Munhall 2006. Participants (n=2, both males) were seated in a soundproof booth and wore an AKG HSC-271 Professional headset. Their speech was routed from the headset microphone through a Delta 44 sound card. Speech was analyzed and resynthesized in realtime. Resynthesized speech was played through the headset's earphones in place of normal auditory feedback.

Before the experiment began, subjects read a short passage to become acclimatized to hearing themselves through headphones. No recording of speech or speech alteration occurred during this period.

Once accustomed to hearing their resynthesized voices, subjects read a set of 7 hVd words off of a computer screen: /hid/, /hɪd/, /hɛd/, /hæd/, /hɑd/, /ho^{w}d/, /hud/. Feedback was not altered during this stage and speech was recorded. Talkers repeated the word list for 20 seconds.

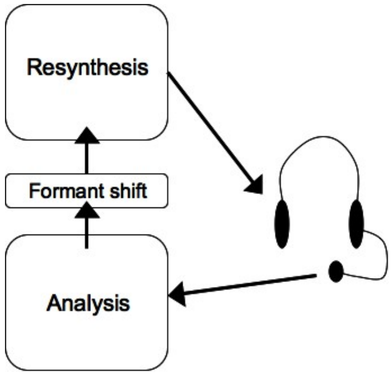

Figure 1: Schematic of Experimental Setup. Subjects speak into the microphone portion of a headset. Their speech is analyzed, then resynthesized (and shifted, if necessary) and fed back into the headset's earphones.

In the alteration stage, a MATLAB program displayed a prompt on the computer screen at regular intervals. To disguise the nature of the task, talkers were told that their reaction time would be recorded as they followed the instructions given by the prompt. They were not informed that their speech would be altered, though a full explanation of the study was given at the end of the experiment.

During each trial, the prompt 'Say HEAD now' was displayed on the computer screen. The first formant of the talker's voice was altered in realtime using a feedback alteration system designed by the second author. The system takes in small time-windows of sound and quickly resynthesizes them using McAulay-Quatieri synthesis, which can replicate a speech signal with very good accuracy from a small amount of data (Quatieri & McAulay, 1986). Post-session interviews indicated that the delay incurred by the analysis and resynthesis, 10ms, is perceptually negligible. Subjects did not notice either formant shifts or delays.

Formant alteration proceeded in four phases:

Phase	# Trials	Formant shift
1	75	0Hz
2	40	0Hz - 200Hz
3	220	200Hz
4	25	0Hz

Initially, formant feedback is left unaltered. After 75 baseline trials, the talker's F1 feedback was slowly raised to 200Hz higher than was actually produced, and remained at the 200Hz maximum shift for 220 trials. Feedback was returned to normal for the last 25 trials.

Results

As the formant values of their auditory feedback were raised, talkers lowered the formant values of the /ɛ/ vowels that they produced. That is, they *compensated* for the change in auditory feedback, closely mirroring the formant patterns observed in previous formant shift experiments. The time course of this effect for a representative subject is illustrated in Figure 2.

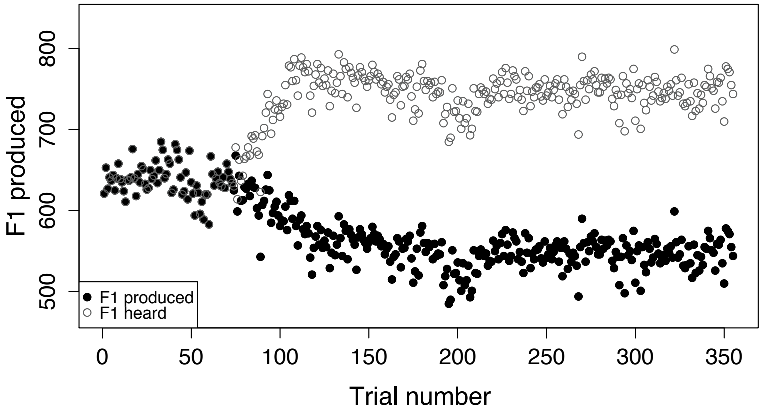

Figure 2: Change in F1 feedback and F1 production in /ɛ/ over the course of the experiment. Open circles indicate the F1 values in talkers' auditory feedback. Filled circles indicate the F1 produced by talkers over the course of the experiment. Each open circle/filled circle pair represents one trial.

Experiment 1

The initial experiment verified that talkers will compensate for changes in their auditory feedback using this experimental setup. To address the relative roles of formants in making these compensatory changes, a modified procedure is necessary.

In Experiment 1, the equipment setup and initial word lists remained the same as in the previous experiment, but the alteration phase differed. In the Experiment 1 alteration phase, formant feedback was slowly raised or lowered to five distinct maximum shifts in a stepwise fashion, resulting in a staircase pattern of formant feedback shifts. Altered feedback remained constant for 20 trials at the top of each "stair". The progression of feedback shifts during the alteration stage is illustrated in Figure 3. There were 7 participants in this experiment, all adult males.

The procedure was repeated once each for F0, F1, and F2 over the course of three days. Order was randomized. One acoustic variable was altered on each of the three days. Only

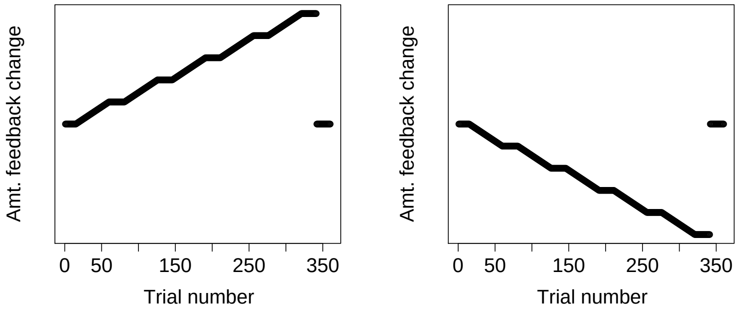

Figure 3: Change in feedback over the course of each experiment. There were 360 trials in each experiment. These 360 trials were composed of 6 regions of equal formant alteration (plateaus) connected by ramps of slowly increasing or decreasing feedback alteration.

the results for shifts in F1 are reported here. Each F1 experiment ramped F1 feedback from baseline (no alteration) to a large maximum alteration of (±150Hz or ±250 Hz).

Maximum changes (in Hertz) for each trial type:

Exp	First Stair	Second Stair	Third Stair	Fourth Stair	Fifth Stair
A	+30 Hz	+60 Hz	+90 Hz	+120 Hz	+150 Hz
B	−30 Hz	−60 Hz	−90 Hz	−120 Hz	−150 Hz
C	+50 Hz	+100 Hz	+150 Hz	+200 Hz	+250 Hz
D	−50 Hz	−100 Hz	−150 Hz	−200 Hz	−250 Hz

Measurements taken

Vowel formants were measured using Esps/xwaves, and verified using PRAAT (Boersma & Weenink, 2008). Degree of compensation (#Hz talker compensation / #Hz feedback alteration) was measured for each trial. For data analysis, the formant values produced by the subject were plotted against the amount of feedback shift. We fit several curves to this data – a line, a logarithmic curve, a negative exponential curve, and a square root curve – and recorded the one with the smallest sum of residuals.

Results

We consider the responses to the largest (250Hz) increases in F1 feedback first, since talkers did not respond as consistently to decreases in F1 feedback shifts or to shifts in F2 feedback. Possible reasons for compensation differences in these conditions are addressed in the discussion.

For F1 feedback increases, there were two major response patterns. In the dominant pattern, exhibited by 5 (of 7) subjects and shown in Figure 4 below, F1 productions continued to decrease until the change in F1 feedback reached approximately 150Hz. Decreases in F1 production for shifts larger than 150Hz were smaller, and in one case, the subject's F1 increased for the largest feedback increases. In all cases, F1 production appears to be approaching a lower bound. For each of these subjects, the negative exponential curve fit best, demonstrating that the amount of compensation for altered F1 feedback decreases as the feedback shift increases.

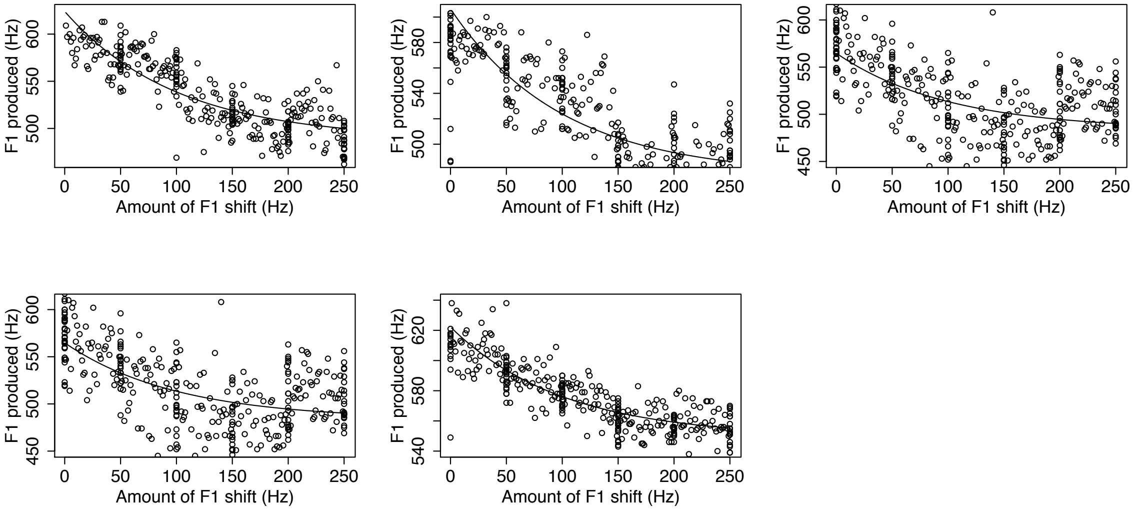

Figure 4: Pattern 1: Asymptotic decrease. Each circle represents one /ɛ/ vowel from one token of 'head'. The best fit negative exponential curve is superimposed. Shift plateaus occurred at 50Hz, 100Hz, 150Hz, 200Hz, and 250Hz. The y axis marks the F1 produced by the subject when their F1 feedback was shifted by the amount indicated on the x axis.

Two subjects exhibited a second response pattern, as shown in Figure 5. Formant production for these subjects continued to decrease through the 250Hz shift in F1 feedback.

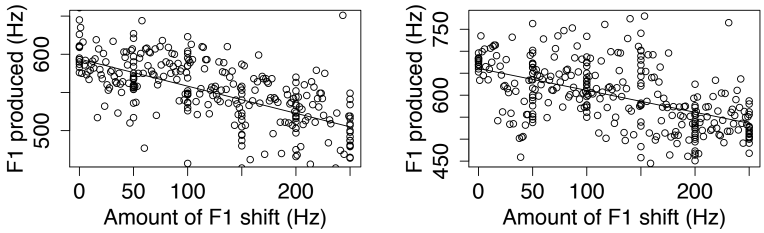

Figure 5: Pattern 2: Linear decrease in F1 production. Each circle represents one /ɛ/ vowel from one token of 'head'. The best fit line is superimposed. Shift plateaus occurred at 50Hz, 100Hz, 150Hz, 200Hz, and 250Hz. The y axis marks the F1 produced by the subject when their F1 feedback was shifted by the amount indicated on the x axis.

The most likely interpretation is that, for most subjects, the percent compensation decreased as the amount of feedback change increased. But because the feedback change increased monotonically over the course of the experiment, it is possible that the decrease in percent compensation was a consequence of time rather than formant change; perhaps most subjects simply lose attention after the first half of the experiment.

Sessions of the same length but with different feedback shift maxima address this concern. If it is the amount of feedback shift rather than the amount of time elapsed that is causing subjects' percent compensation to decrease for feedback shifts greater than 150Hz, we would expect that formant production in subjects with maximum shifts of 150Hz or less should resemble the left half of the nonlinear response graphs above. Subjects with smaller maximum feedback shifts should show a steady linear decrease in their formant production.

Experiment 2

Method

In this experiment, the maximum shift is reduced from 250Hz to 90Hz, well below the compensation asymptote. Because the maximum auditory feedback shifts were much smaller in this experiment, the number of stairs was lowered from 5 to 2, and the number of trials at each stair was raised from 20 to 90. Stairs were located at +30Hz and +90Hz. The procedure was otherwise identical to Experiment 1. The 5 participants in this experiment were all adult males.

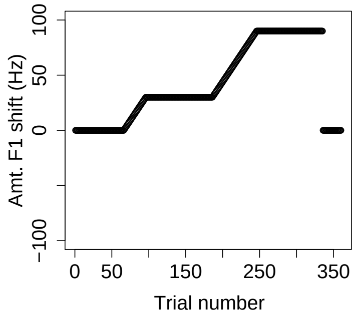

Figure 6: Change in auditory feedback in the small feedback shift condition. There were 360 trials in the experiment and two formant shift plateaus, at 30Hz and 90Hz. Each shift plateau lasted for 90 trials. Feedback returned to normal at the end of the experiment.

Results

As expected, the dominant response reverses for smaller shifts in auditory feedback. Whereas increases in feedback shifts up to 250Hz result in compensation increasing toward an asymptote, smaller increases display a more linear pattern. As illustrated by Figure 7, compensation responses from four of the five subjects were a better fit for a linear function.

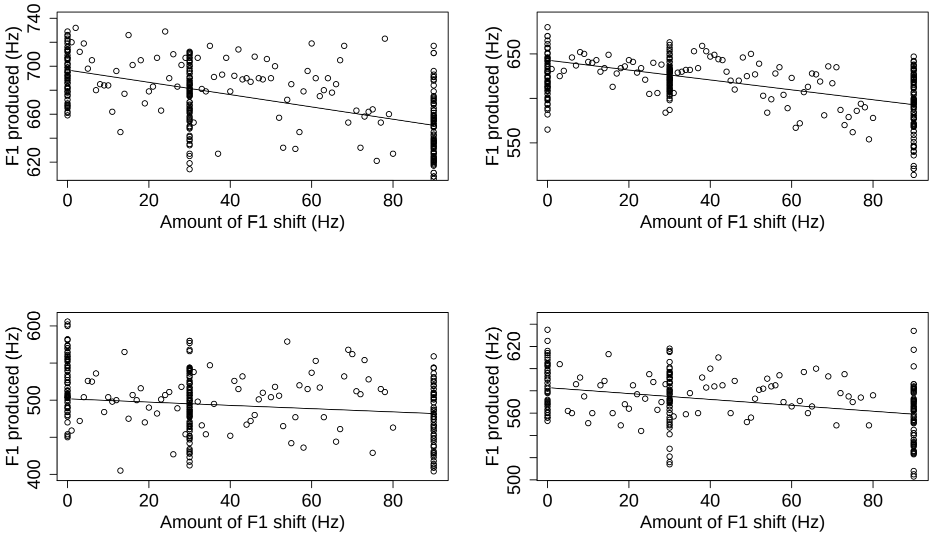

Figure 7: Pattern 1: Linear decrease in F1 production. Each circle represents one token of 'head'. The best fit line is superimposed. The four subjects followed feedback shift pattern E, with feedback shift "stairs" at 30Hz and 90Hz. The y axis marks the F1 produced by the subject when their F1 feedback was shifted by the amount indicated on the x axis.

Only one subject's data had a better nonlinear fit, shown in Figure 8.

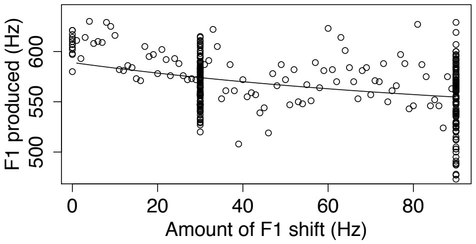

Figure 8: Pattern 2: Asymptotic decrease. Each circle represents one token of 'head'. The best fit negative exponential curve is superimposed. This subject followed feedback shift pattern E, with feedback shift "stairs" at 30Hz and 90Hz. The y axis marks the F1 produced by the subject when their F1 feedback was shifted by the amount indicated on the x axis.

Intermission

Experiments 1 and 2 have shown that talkers compensate even for small shifts in auditory feedback. However, as in previous feedback shift experiments, compensation in these experiments was incomplete. If compensation were complete, we would expect a perfect correspondence between F1 feedback shifts and F1 productions. When F1 feedback is raised by 50Hz, talkers should lower the F1 they produce by 50Hz. Thus in Experiment 1, where F1 feedback shifts ranged from 0Hz to 250Hz, we would expect the range in F1 productions to span approximately 250Hz. In fact, the F1 production range is about half of this amount. If compensation were complete in Experiment 1, we would also expect compensation to be linear with feedback shift. In fact, compensation in the majority of talkers was nonlinear, with less compensation for large feedback shifts.

In Experiment 2, the auditory feedback shifts were smaller, and talker compensation increased linearly with the amount of feedback shift. Taken together, these results suggest that compensation increases linearly with feedback shift up to a compensation limit, and that talkers are reluctant to produce formants beyond this limit. Because /ɛ/ is a mid-vowel, subjects are clearly able to reduce their F1 further. Why don't they?

Experiment 3

Experiments 1 and 2 have shown that talkers are sensitive to even small shifts in auditory feedback, and that compensation is not complete. Determining the completeness of compensation requires assuming a single baseline value to F1 for /ɛ/, an assumption that seems unreasonable given that there are 360 /ɛ/ productions to consider.

Either there is a single value of F1 that is considered in the talker's /ɛ/ production, or somatosensory feedback contributes to the decision of how much to compensate. In the latter case, completeness of compensation should decrease as talkers leave their baseline regions for /ɛ/. When they produce vowels that, articulatorily, are no longer productions of /ɛ/, their compensation should decrease.

Experiment 3 distinguished between these two possibilities. The experimental procedure was identical to the procedure described above, but left voice feedback unaltered. All participants in Experiment 1 and Experiment 2 also took part in Experiment 3. Talkers thus served as their own controls.

Over 360 trials, talkers produced a variety of /ɛ/ formants. The standard deviation of vowels ranged from 15-30Hz, consistent with previous literature (Purcell & Munhall, 2006). Results from a typical subject are shown below, alone and with his Experiment 1 vowels superimposed.

This subject's control vowels from Experiment 3 are shown in F1-F2 space in the plot on the left. Over 360 trials, the subject covers a significant portion of this space, nearly intersecting with his average 'had' production. In the graph on the right, we can see that altered feedback causes this subject to produce vowels outside of his prototypical /ɛ/ region.

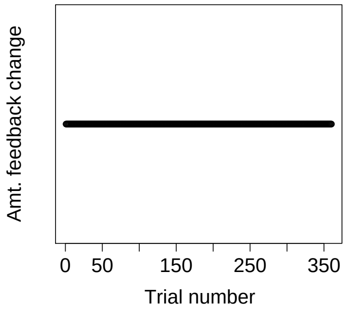

Figure 9: Change in auditory feedback in the control condition. There were 360 trials in the experiment, during which auditory feedback remained unaltered.

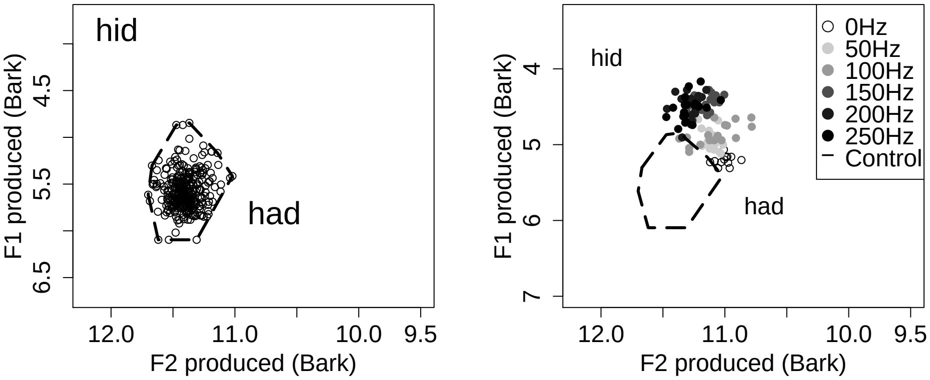

Figure 10: (a) Productions of /ɛ/ during control trials for a typical subject (here, subject 4). Dots represent individual vowel tokens plotted in F1-F2 space. Surrounding shape is a convex hull outlining baseline vowel tokens. (b) Convex hull of control /ɛ/ productions superimposed on /ɛ/ productions during the F1 test trials. Darker dots are vowels produced with larger increases in F1 feedback. Clearly these darker dots are produced with lowered F1, visible as dots *higher* on the y axis.

This was a general pattern. When compensating, subjects are willing to tolerate potentially abnormal somatosensory feedback in the service of righting auditory feedback. This does not mean, however, that they are willing to tolerate articulations that are arbitrarily far from their normal range.

It is still possible that abnormal somatosensory feedback is responsible for the decreasing completeness in compensation as auditory feedback shift increases. When compensation is complete, the talker's vowels should fall within his control /ɛ/ region even after they are shifted. We can evaluate the completeness of compensation by looking at

1. how far vowels that talkers produce stray from the control vowel space
2. how far vowels that talkers hear (≡ vowels that talkers produce after they are shifted) stray from the control vowel space

Below, the vowels that each subject produced and heard in Experiment 1 are plotted against the control vowels from Experiment 3.

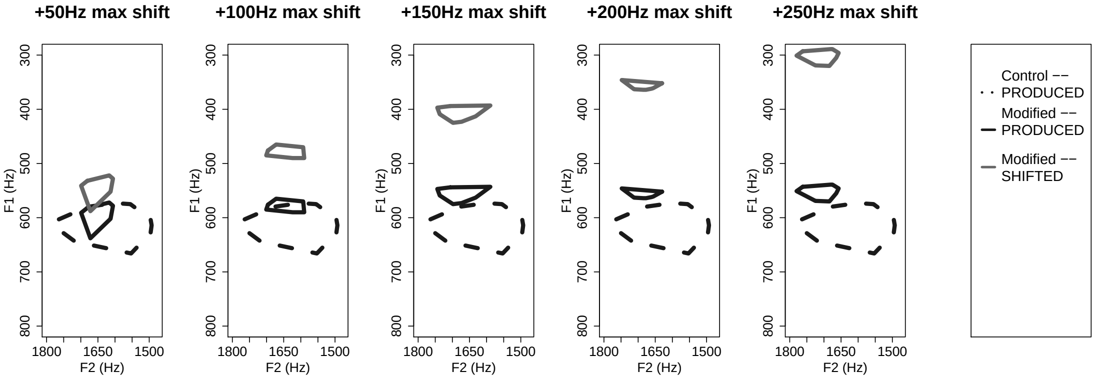

Figure 11: Productions of /ɛ/ during Experiment 1 plotted in F1-F2 space against productions of /ɛ/ during control trials. Results for a typical subject are shown (here, subject 7). For small feedback shifts, the solid black shape falls entirely within the dotted black shape, indicating that the vowels that the subject heard were all within their baseline region and compensation was complete. As the amount of feedback shift increases, compensation is less and less complete. Further explanation is given in the body of the paper.

Each of the graphs in Figure 11 shows the /ɛ/ vowels produced at each of the five F1 feedback shift maxima – 50Hz, 100Hz, 150Hz, 200Hz, and 250Hz. The dotted shape in each graph outlines the vowels produced during control trials. The solid gray shape in each graph represents the convex hull of vowels produced with F1 feedback shifted by the amount shown in the graph title. For example, the solid gray shape in the leftmost graph outlines the vowels produced during trials with F1 feedback shifted by +50Hz. The solid black shape in each graph outlines these vowels after they have been shifted by 50Hz. These are the vowels that subjects heard. For example, a vowel produced with an F1 of 550Hz in the leftmost graph would fall within the gray shape. That same vowel, after its 50Hz shift, would be located at 600Hz and would fall in the solid black shape. This latter vowel is what the subject heard. Notice that in the leftmost graph, the solid black shape falls completely within the dotted black line, indicating that the shifted vowels that the subject heard were all within his baseline region, and that compensation was complete. Compensation is likewise nearly complete for F1 feedback shifts of 100Hz. As the amount of feedback shift increases to 150Hz and beyond, the vowels that the subject hears are no longer within his baseline region and compensation is less and less complete.

Discussion

Taken together, Experiments 1-3 suggest that both auditory and somatosensory feedback are essential components of speech planning. Experiment 1 showed that abnormal auditory feedback causes talkers to compensate, and the compensation was greater for small shifts in F1 feedback than for large shifts in feedback. Experiment 2 confirmed this observation by showing that small stepwise shifts in F1 feedback usually caused linear compensation for shifts in auditory feedback. Experiment 3 showed that compensation was complete for small shifts in auditory feedback and partial for large shifts in auditory feedback.

Note that the nonlinearity of compensation observed in Experiment 1, which started at feedback shifts of 150Hz, coincides with productions falling outside of the baseline region. We suggest that the decreasing completeness in compensation arises from the juxtaposition of abnormal auditory feedback and normal somatosensory feedback. For small shifts in F1 feedback, auditory feedback is abnormal and somatosensory feedback is normal. There is only a small discrepancy between the two types of feedback, and subjects are able to ignore the normal somatosensory feedback in order to compensate for the acoustic feedback. For large shifts in F1 feedback, auditory feedback is highly abnormal, but somatosensory feedback remains normal. There is a huge discrepancy between the two. Because both types of feedback are taken into account when planning the next utterance, the somatosensory signal attenuates the compensation response. Talkers will compensate, but they have a limit; they will not produce vowels that fall too far outside of their baseline regions.

These results are consistent with models of planning which use contributions from both auditory and somatosensory feedback. Critically, the relative contributions of these two sources varies with their distance from baseline. For small shifts in auditory feedback, auditory feedback seems to swamp normalizing somatosensory feedback. For larger shifts in auditory feedback, somatosensory feedback attenuates the compensation response. The interplay of these two sources of feedback stabilizes the speech production system. It is also notable that there are large individual differences in degree of compensation, suggesting that the amount of compensation driven by abnormal feedback differs between talkers.

The results of these experiments have close ties to perturbation experiments involving the Lombard effect, in which talkers compensate for changes in the perceived loudness of their voices. The profile of these responses parallel responses to our pitch changes closely. When loudness is perturbed by a small amount, less than 1dB, compensation is nearly complete. Large perturbations result in much smaller proportional changes (Heinks-Moldenado & Houde, 2005; Bauer, Mittal, Larson, & Hain, 2006). Results were similar in spite of the fact that these were somewhat less linguistic tasks (subjects in these studies held out the vowel /u/ rather than saying real words).

Given these two results, we would predict that compensation would be stronger if conflicting somatosensory feedback were eliminated. A recent experiment by Larson *et al* tests this hypothesis for pitch shifts (Larson et al., 2008). In this experiment, talkers produced the vowel /u/ before and after their vocal cords were numbed by anesthesia. There was a significant difference in degree of compensation before and after the anesthesia was administered: beforehand, subjects produced an average of 20 cents compensation for a 100 cent change in auditory feedback, whereas after the anesthesia was administered, subjects compensated for a 100 cent feedback change by about 25 Hz. Larson finds two possible interpretations for these results. First, somatosensory feedback might oppose auditory feedback. Second, the lack of somatosensory feedback might trigger the monitoring system to attend more to auditory feedback.

We can now return to our four hypotheses for partial compensation. Fuzzy targets are not a good explanation for the results of Experiments 1-3 because compensation consistently opposes the change in feedback, and the pattern of compensation across shifts in F1 is fairly consistent across participants. If targets were imprecise, we would see a wide range of unsystematic responses to changes in auditory feedback. Bone conduction is also an unlikely explanation for these results because early feedback studies (e.g. Houde & Jordan, 1998) used whispered speech, which is not confounded by bone conduction, and still found partial compensation.

Two explanations remain: either feedback is a function of F1 (or of multiple formants), or both auditory and somatosensory feedback is incorporated into the target. We have just explained why feedback from both types of sources is a good fit for the data collected in these three experiments. However, we should consider whether the nonlinear compensation that we see in F1 might be an artifact of linear compensation for some unknown auditory feature that varies nonlinearly with F1. In particular, a convex function of F1 could yield the compensation patterns observed in experiment 1; if talkers track some attribute which increases as F1 increases, but at a decreasing rate, we would see diminishing compensation as we keep changing F1. This could be true, for example, if talkers were tracking the ratio of F2/F1. This sort of explanation – or language-specific differences in perception – might also account for the asymmetry in compensation profiles between F1 and F2. These are worthwhile questions to be taken up in future work.

Whether we are storing F1 or some function correlated with F1, we argue that a purely auditory target is unlikely given that changes in somatosensory feedback alone generate compensatory responses (Tremblay et al., 2003), and that compensation for shifts in pitch feedback changes when somatosensory feedback is removed (Larson et al., 2008).

Conclusions

Speech monitoring and planning incorporates both auditory and somatosensory feedback. For small discrepancies between auditory and somatosensory feedback, auditory feedback takes precedence. For large discrepancies between auditory and somatosensory feedback, somatosensory feedback takes precedence.

References

Bauer, J. J., Mittal, J., Larson, C. R., & Hain, T. C. (2006). Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude. *Journal of the Acoustical Society of America*, *119*(4), 2363-2761.

Boersma, P., & Weenink, D. (2008). Praat: doing phonetics by computer (version 5.0.08) [computer program]. *http://www.praat.org/.*

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice f0 responses to manipulations in pitch feedback. *Journal of the Acoustical Society of America*, *103*(6), 3153-3161.

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, *102*(3), 594-621.

Guenther, F. H., & Barreca, D. M. (1997). Self organization, computational maps and motor control. In P. Morasso & V. Sanguineti (Eds.), (p. 383-421). Amsterdam, North Holland.

Heinks-Moldenado, T. H., & Houde, J. F. (2005). Compensatory responses to brief perturbations of speech amplitude. *ARLO*, *6*(3), 131-137.

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

Houde, J. F., & Jordan, M. I. (2002). Sensorimotor adaptation of speech i: Compensation and adaptation. *Journal of Speech, Language, and Hearing Research*, *45*(2), 295-310.

Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of f0 production: Evidence from feedback perturbation. *Journal of the Acoustical Society of America*, *108*(3), 1246-1251.

Jones, J. A., & Munhall, K. G. (2005). Remapping auditory-motor representations in voice production. *Current Biology*, *15*, 1768-1772.

Larson, C. R., Altman, K. W., Liu, H., & Hain, T. C. (2008). Interactions between auditory and somatosensory feedback for voice f0 control. *Experimental Brain Research*, *187*, 613-621.

Pascal Perrier, H. L., & Payan, Y. (1996). Control of tongue movements in speech: the equilibrium point hypothesis perspective. *Journal of Phonetics*, *24*(53-75).

Pile, E. J. S., Dajani, H. R., Purcell, D. W., & Munhall, K. G. (2007). *Talking under conditions of altered auditory feedback: Does adaptation of one vowel generalize to other vowels?*

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

Quatieri, T. F., & McAulay, R. J. (1986). Speech transformations based on a sinusoidal representation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *ASSP-34*(4), 1449-1464.

Sanguineti, V., Laboissière, R., & Ostry, D. J. (1998). A dynamic biomechanical model for neural control of speech production. *Journal of the Acoustical Society of America*, *103*(3), 1615-1627.

Schroeder, C. E., Lindsley, R. W., Specht, C., Marcovici, A., Smiley, J. F., & Javitt, D. C. (2001). Somatosensory input to auditory association cortex in the macaque monkey. *The Journal of Neurophysiology*, *85*(3), 1322-1327.

Shiba, K., Miura, T., Yuza, J., Sakamoto, T., & Nakajima, Y. (1999). Laryngeal afferent inputs during vocalization in the cat. *NeuroReport*, *10*(5), 987-991.

Tremblay, S., Shiller, D. M., & Ostry, D. J. (2003). Somatosensory basis of speech production. *Nature*, *423*, 866-869.

Wyke, B. (1983). Vocal fold physiology: Contemporary research and clinical issues. In D. M. Bless & J. H. Abbs (Eds.), (chap. Neuromuscular Control Systems in Voice Production). San Diego: College Hill Press.