Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate

K. G. Munhall[a)]

Department of Psychology and Department of Otolaryngology, Queen's University, Humphrey Hall, 62 Arch Street, Kingston, Ontario K7L 3N6, Canada

E. N. MacDonald

Department of Psychology and Department of Electrical and Computer Engineering, Queen's University, Humphrey Hall, 62 Arch Street, Kingston, Ontario K7L 3N6, Canada

S. K. Byrne and I. Johnsrude

Department of Psychology Queen's University, Humphrey Hall, 62 Arch Street, Kingston, Ontario K7L 3N6, Canada

Talkers show sensitivity to a range of perturbations of auditory feedback (e.g., manipulation of vocal amplitude, fundamental frequency and formant frequency). Here, 50 subjects spoke a monosyllable ("head"), and the formants in their speech were shifted in real time using a custom signal processing system that provided feedback over headphones. First and second formants were altered so that the auditory feedback matched subjects' production of "had." Three different instructions were tested: (1) *control*, in which subjects were naïve about the feedback manipulation, (2) *ignore headphones*, in which subjects were told that their voice might sound different and to ignore what they heard in the headphones, and (3) *avoid compensation*, in which subjects were informed in detail about the manipulation and were told not to compensate. Despite explicit instruction to ignore the feedback changes, subjects produced a robust compensation in all conditions. There were no differences in the magnitudes of the first or second formant changes between groups. In general, subjects altered their vowel formant values in a direction opposite to the perturbation, as if to cancel its effects. These results suggest that compensation in the face of formant perturbation is relatively automatic, and the response is not easily modified by conscious strategy.
© 2009 Acoustical Society of America. [DOI: 10.1121/1.3035829]

I. INTRODUCTION

Human speech and animal vocalizations are dramatically influenced by the sounds that the speakers hear themselves producing (Smotherman, 2007). Both clinical and laboratory studies demonstrate that this auditory feedback effect occurs because vocal motor control is normally dependent on the sensory consequences of talking. Hearing-impaired individuals show characteristic patterns of distortion and increased speech variability in the absence of normal feedback (Cowie and Douglas-Cowie, 1992), and a range of experimental perturbations of acoustic feedback produce rapid compensations in subsequent productions (Burnett *et al.*, 1998; Houde and Jordan, 1998; Kawahara, 1995; Lane and Tranel, 1971; Purcell and Munhall, 2006; Villacorta *et al.*, 2007). Similar phenomena can be demonstrated in a variety of species ranging from songbirds (Brainard and Doupe, 2000) to beluga whales (Scheifele *et al.*, 2005).

One of the most common questions about the way in which auditory feedback affects speech production is whether subjects are aware that the feedback is being ma-nipulated. An implication of this concern is that subjects' compensations might be under conscious control or result from a response strategy. This is a complex question that invokes an old controversy in the neuroscience of behavior—the idea of a reflex and the conflict between the ideas of voluntary or consciously controlled action and involuntary or automatic movements.

This controversy between volitional and automatic control is at the core of both philosophical and neurophysiological debates about the control of action [see Prochazka *et al.* (2000) for a discussion], and its resolution is beyond the scope of this paper. However, empirical contributions to the question of voluntary/involuntary responses can make the discussion more explicit and well defined. Three types of data bear on the reflexive nature of response to sensory stimulation. First, one can investigate the timing of motor responses: more rapid responses can be viewed as more automatic. Second, the influence of training can be investigated in order to determine whether a response is modifiable. Finally, the influence of instructions or task on the stereotypy of the response can be assessed. If the response is unchanged by training, instructions, or task, it is more likely to be automatic and to operate independently of conscious control.

Each of these general approaches is evident in the sensorimotor control of speech literature. Rapid motor responses

---

a)Author to whom correspondence should be addressed. Electronic mail: kevin.munhall@queensu.ca

to mechanical perturbations have been frequently reported [e.g., see Gracco and Abbs (1985)], and the responses exhibit distinct patterns depending on the timing of the perturbation. Responses to auditory feedback perturbations also depend on the time from the onset of the perturbation. Rapid responses to perturbations of vocal amplitude (Bauer *et al.*, 2006) and vocal pitch (Burnett *et al.*, 1998; Hain *et al.*, 2000) have been reported, but a later response can be observed in some cases (Hain *et al.*, 2000; Kawahara, 1995). A distinction between automatic rapid responses (<200 ms) and a more voluntary slower response between 300 and 700 ms after perturbation onset has been suggested [e.g., see Hain *et al.* (2000)].

When subjects were given experience with perturbed auditory feedback, there is some evidence of an increased independence from the auditory feedback signal and thus an ability to resist compensatory behavior. Pick *et al.* (1989) exposed subjects to an increased noise level with visual feedback about their speaking level. This visual feedback was effective in helping people resist the effects of background noise level, but their learning seemed to be associated with a strategy to reduce the overall speaking level under all auditory conditions. Zarate and Zatorre (2008) found that compared to nonmusicians, trained singers are much better able to ignore pitch perturbations. However, even the singers made small compensations in response to the voice feedback pitch shifts.

Instructions or task orientation has been found to produce limited modifications of responses under certain conditions. Pick *et al.* (1989) found that there was little influence of instructions to ignore the presence of background noise. Their subjects exhibited the Lombard effect and increased their vocal amplitude to match the background noise even when they had been explicitly instructed not to. With the same instructions, Hain *et al.* (2000) found that subjects always produced changes in vocal pitch when fundamental frequency was perturbed. However, when instructed to raise or lower their pitch in response to a perturbation or move in the opposite direction to the perturbation, subjects did show the ability to make changes in the timing and magnitudes of compensation.

In the present study, we focus on the response to perturbations of vowel formant frequency. Rapid signal processing systems now allow the frequency of one or more formants to be shifted up or down in frequency in real time. In response to changes in auditory feedback, talkers adjust the frequency of their produced formants in the opposite direction in frequency, presumably in order to compensate for the perturbations (Houde and Jordan, 1998; Purcell and Munhall, 2006; Villacorta *et al.*, 2007). These compensations persist when feedback is returned to normal, suggesting that some type of learning has taken place. However, the extent to which such compensations are relatively automatic is uncertain.

Many of the initial studies of speech compensation used a very gradual perturbation in which the discrepancy between the actual formant frequency and the modified feedback was changed by small increments trial by trial. The subjects in these studies often seem to be unaware that their speech was being modified and thus have no particular knowledge of the nature of the manipulation [e.g., see Pur-

cell and Munhall (2006)]. However, the same feedback perturbations can be carried out more abruptly by changing the formant frequencies in larger increments (step changes). These abrupt step changes can be noticeable and thus introduce the possibility of a more explicit strategic response.

Here, three groups of subjects are tested under different instructional conditions using this sudden large perturbation paradigm: (1) subjects who are naive to the purposes of the experiment and are not told about the feedback perturbation, (2) subjects who are told that their speech heard from the headphones may sound wrong and that they should ignore this feedback, and (3) subjects who are briefed in detail about the perturbation paradigm and instructed to not compensate for the perturbation. The aim is to examine the influence of such instructions on the pattern of formant compensation. If the second or third instruction condition reduces or eliminates the compensatory response, then we may conclude that the role of auditory feedback in speech motor control is not mandatory and is instead open to cognitive intervention. If, on the other hand, the same pattern of response is evident across all three instructional conditions, this would suggest that the maintenance of formant frequency is a more automatic response customarily tuned by auditory feedback. If so, we may conclude that studies of formant perturbation, even when the changes are large and abrupt, are minimally influenced by strategic efforts of subjects since the standard response is difficult to suppress.

II. METHODS

A. Subjects

Fifty-four female participants (mean age=20.1 yr, range: 17–25 yr) were tested in a single session. Since our experiment involves tracking individual formants, we chose to test females exclusively in order to minimize the variability in frequency of the first and second formants across participants. All subjects spoke English as their first language and reported no speech or language impairments. Hearing thresholds were assessed over a range of 500–4000 Hz Three subjects were eliminated because of heightened hearing thresholds in some frequency bands (>20 dB HL). Data from one additional subject were lost due to experimenter error.

B. Equipment and real-time formant shifting

Equipment was the same as that previously reported in Purcell and Munhall (2006). Subjects' speech was recorded using a headset microphone (Shure WH20). The signal was amplified using a Tucker-Davis Technologies MA3 microphone amplifier and low-pass filtered at a cutoff frequency of 4500 Hz (Frequency Devices 901 filter). This signal was digitized at a 10 kHz sampling rate and was filtered in real time to produce formant shifts using a National Instruments PXI-8176 controller. Noise was added using a Madsen Midimate 622 audiometer, and the voice signal and noise were presented to the subject using headphones (Sennheiser HD 265).

Detection of voicing and shifting of formants was performed as previously described in Purcell and Munhall

J. Acoust. Soc. Am., Vol. 125, No. 1, January 2009

Munhall *et al.*: Instructions and response to formant perturbation 385

(2006). Briefly, the manipulation of auditory feedback was achieved by filtering the voice in real time. Voicing was detected using a statistical amplitude threshold technique. Formants in the speech were determined using an iterative Burg algorithm (Orfandidis, 1988). The formant estimates were used to calculate the filter coefficients so that a pair of spectral zeroes was positioned at the location of the existing formant frequency and a pair of spectral poles was positioned at the desired frequency of the new formant. This filtering reduced the spectral energy in the region of the produced formants and emphasized the energy in the region of the desired formants. The filtering and thus the formant shifts were implemented as soon as voicing was detected. The formant frequency estimate and new filter coefficients were computed every 900 μs.

C. Procedure and experimental conditions

Testing was performed in an Industrial Acoustics Company sound insulated room. Prior to data collection, a screening procedure was carried out to determine the best autoregressive model order for formant tracking. Subjects produced seven English vowels spaced across the vowel space in an /hVd/ context five times in a random order. During the experiment subjects produced 95 repetitions of "head" at a natural rate and speaking level with timing controlled by a visual prompt on a monitor.

The subjects were randomly assigned to one of three conditions in which the experimental instructions were manipulated: (1) *control* ($N=18$), in which subjects were naïve about the feedback manipulation; (2) *ignore headphones* ($N=15$), in which subjects were told that their voice might sound different and to ignore what they heard in the headphones; (3) *avoid compensation* ($N=17$), in which subjects were informed in detail about the manipulation and were told not to compensate. This group was given the suggestion that focusing on their kinesthetic feedback might help them avoid compensation. All three groups produced the English word "head" repeatedly in the following experimental phases (see Fig. 1): (1) *baseline*. Fifteen repetitions were spoken with normal feedback (i.e., amplified and with noise added but no shift in formant frequency) to assess baseline F1 and F2 values. In this and subsequent conditions, subjects were encouraged to speak at a natural rate and speaking level with timing controlled by a prompt on a monitor. Each prompt lasted 2.5 s, and the intertrial interval was approximately 1.5 s. (2) *Perturbation*. Forty repetitions of the utterance "head" were produced with F1 and F2 values shifted in frequency to match the formant values for each subject's production of the vowel /æ/ as in "had." (3) *Return to normal feedback*. Forty repetitions of the utterance were produced with normal feedback (i.e., the formant shift was abruptly turned off).

D. Offline formant analysis

In order to examine the extent to which the shifting of formants affected the acoustics of produced vowels on subsequent trials, segmentation boundaries for the vowel in each trial were first calculated using an automated process that examined the harmonicity of the power spectrum. These
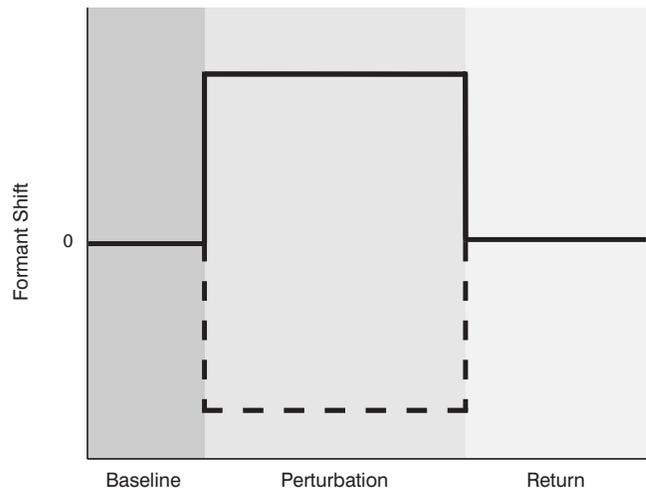

FIG. 1. Feedback shift for the three phases of the experiment. In the baseline condition (15 utterances), the feedback was normal (unperturbed). For the perturbed condition (40 utterances), the first formant was increased (solid line) and the second formant was decreased (dashed line). The size of the shift for each formant was calculated as the difference between the average frequency for that formant in "head" and in "had" separately for each individual. For the return condition (40 trials), the feedback was returned to normal.

boundaries were then inspected visually and corrected if required. Once segmented, offline estimates of the formant frequencies were calculated by sliding a 25 ms analysis window for each estimate, one ms at a time (shifts of ten speech samples), and applying a similar algorithm to that used in online shifting. For each trial, a single "steady-state" F1 value was determined by averaging 40% of the F1 estimates starting almost halfway through the vowel (i.e., from 40% of the way to 80% of the way through the vowel). Single steady-state values for F2 and F3 were calculated in the same manner. Prior to data collection, a screening procedure was conducted to select the best autoregressive model order (a parameter used in the real-time formant tracking algorithm) for each talker. This reduced gross errors in formant tracking. However, for some participants, occasionally, one of the formants would be misinterpreted as another (e.g., F2 being misinterpreted as F1). These misinterpreted estimates were found and corrected by visually examining a plot with all of the steady-state F1, F2, and F3 estimates for each individual.

III. RESULTS

While the frequency shifts applied to F1 and F2 in the perturbed condition varied for each individual, the average frequencies of the feedback shifts were similar across the three instruction groups. This was confirmed by an analysis of variance (ANOVA) with no significant main effect of instruction group for either the F1 [$F(2,47)=0.696$, $p=0.50$] or the F2 [$F(2,47)=0.115$, $p=0.89$] frequency shift. See Table I for the mean perturbed feedback shifts.

In order to determine the magnitude of subjects' compensations to these shifts, the formant estimates for their produced vowels were normalized for each individual by subtracting their baseline mean, defined as the mean of the estimates for trials 6–15. These trials correspond to the last ten utterances in the baseline condition. For each trial, the

Munhall *et al.*: Instructions and response to formant perturbation

TABLE I. Mean formant feedback shift and compensation in Hz for F1 and F2 for the three instruction conditions. Standard errors of the means are shown in parentheses.

	Control	Ignore headphones	Avoid compensation
F1 shift	183.2(15.7)	209.2(26.3)	221.1(28.6)
F1 compensation	−56.6(6.0)	−61.0(13.7)	−55.2(5.8)
F2 shift	−235.2(18.8)	−216.8(28.7)	−224.9(33.0)
F2 compensation	68.3(11.3)	54.9(14.6)	55.0(16.2)

TABLE II. Mean formant frequency in Hz for the baseline, perturbation, and return phases for the three instruction conditions. Standard errors of the means are shown in parentheses.

Conditions		F1	F2
Baseline:	Avoid	740.0(13.6)	2078.1(46.4)
	Ignore	732.2(18.1)	2071.6(31.8)
	Control	762.3(10.6)	2062.2(25.8)
Perturbation:	Avoid	686.4(10.9)	2126.4(50.7)
	Ignore	671.5(17.9)	2127.0(38.0)
	Control	711.2(9.5)	2122.4(21.7)
Return:	Avoid	723.6(15.2)	2088.6(48.6)
	Ignore	733.2(15.2)	2053.5(33.5)
	Control	750.8(9.7)	2062.3(25.0)

normalized F1 and F2 estimates were averaged for each group and are plotted in Fig. 2. From the figure, it is clear that on average all three groups changed their production of F1 and F2 in a direction opposite to the manipulation.

To quantify the change in production, three intervals were defined based on the last ten trials in each of the experimental conditions (trials 6–15 for the Baseline, trials 46–55 for the perturbation phase, and trials 86–95 for the return phase). In these three intervals, it is assumed that formant production has reached a steady state. The non-normalized F1 and F2 estimates in each interval were averaged for each individual (Table II). A repeated measure ANOVA with interval as a within-subjects factor and instruction group as a between-subjects factor confirmed a significant effect of interval for both F1 $[F(1.7,94)=76.52, p<0.001]$ and F2 $[F(2,94)=32.60, p<0.001]$. Multiple pairwise comparisons using Bonferroni correction confirmed that the results for the perturbation phase were significantly different from both the baseline and return phases. (See Table I for the mean compensation magnitude computed by subtracting the baseline mean from the perturbation mean for each subject.) The difference between the baseline and return phases was not significant. No significant effect of instruction group was found for either F1 $[F(2,47)=1.61, p=0.21]$ or F2 $[F(2,47)=0.06, p=0.95]$. Also, no significant interaction of interval and instruction group was found for either F1

$[F(3.4,94)=1.33, p=0.27]$ or F2 $[F(4,94)=0.96, p=0.43]$. This confirms that the instructional set did not affect the magnitude of the compensation in either F1 or F2; subjects modified their productions even when instructed explicitly not to.

Although the averaged results presented in Fig. 2 show consistent compensation, it is important to note that the individual responses vary greatly. The compensation in both F1 and F2 of each individual, defined by measuring the difference in the average formant frequency between the perturbation and baseline intervals used above, is plotted in Fig. 3. From the figure, it is clear that there is a wide range of compensation across individuals in both F1 and F2 for all three groups. For each of the three instruction conditions, there was also at least one subject who compensated by changing their formants in the direction of the perturbation. Following behavior similar to this has been observed in other auditory perturbation experiments [e.g., Burnett et al. (1998)]. Despite this large intersubject variability, it is apparent that there is consistent compensation for most subjects and a broad similarity across conditions. We also observe a
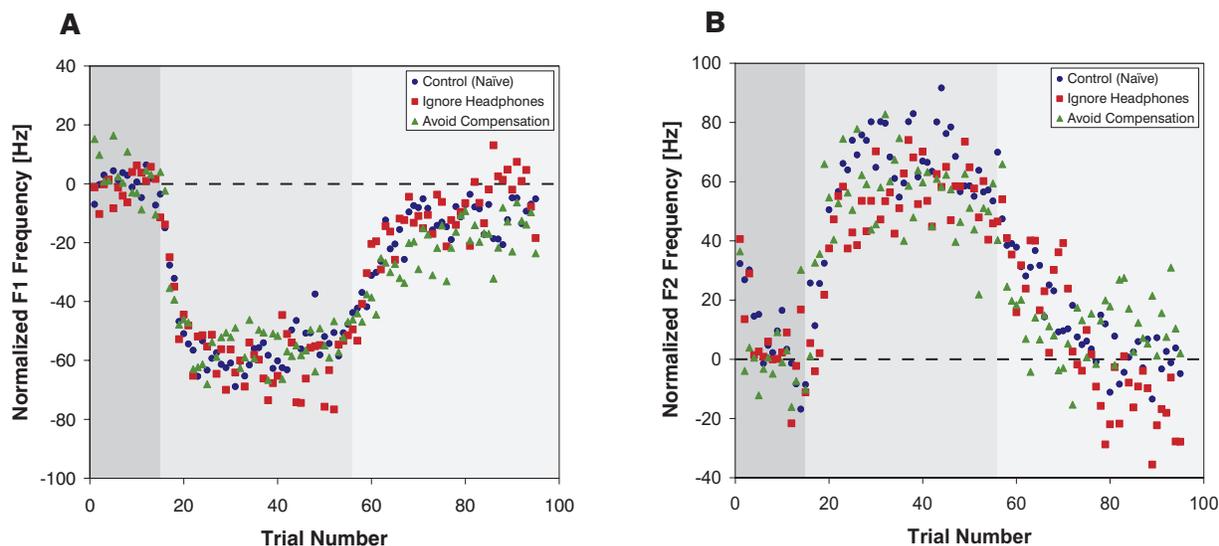

FIG. 2. Average normalized F1 and F2 frequencies for each trial for the control (circles), ignore headphones (squares), and avoid compensation (triangles) groups. (a) and (b) show the results in each trial for F1 and F2, respectively. The three phases of the experiment (baseline, perturbed, and return) are indicated with shading of increasing lightness.
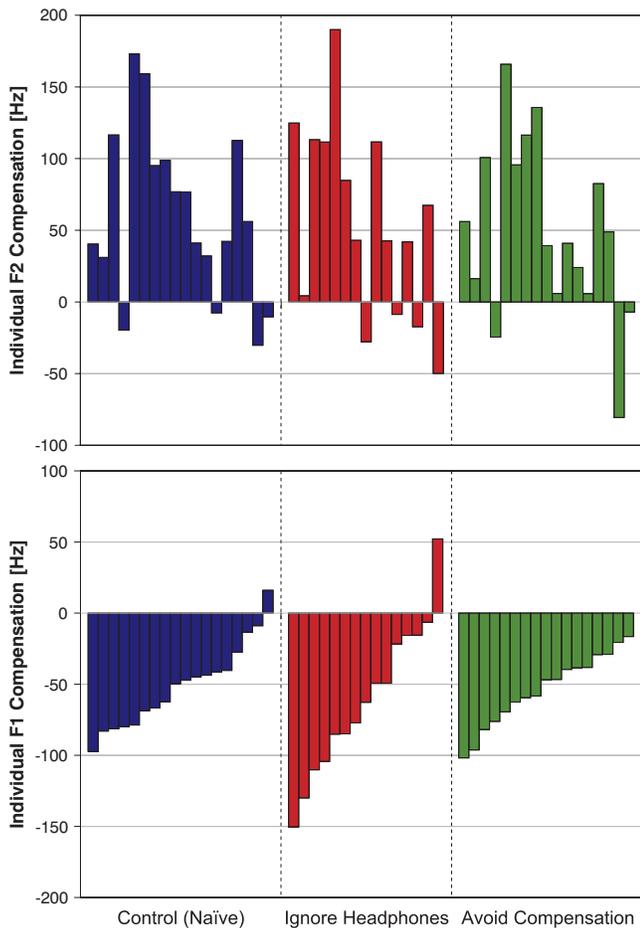
FIG. 3. Individual compensation magnitudes measured as the difference in the average formant frequency between the perturbation and baseline intervals for F2 (top panel) and F1 (bottom panel) for subjects in the control ($N=18$), the ignore headphones ($N=15$), and the avoid compensation ($N=17$) groups. Each bar represents a different subject; homologous bars in the top and bottom panels are from the same subject.

small correlation between the magnitudes of compensation in F1 and F2, $r(48)=-0.44$ and $p<0.001$. This correlation suggests that a common underlying factor might influence compensation for perturbation of both formants.

IV. DISCUSSION

Robust and consistent compensations were observed in all instruction conditions when the first two formants of auditory feedback during speech production were perturbed in real time. The perturbations were individually determined and modified the feedback from the characteristic formant frequencies for each individual's /ɛ/ vowel to the formant values normally expected for their own /æ/ vowel. To compensate, subjects almost always produced formants that were shifted in the opposite direction from the perturbation. F1 values were shifted downward in response to a perturbation that raised the frequency, whereas the F2 frequency was increased in response to a perturbation that lowered the F2 feedback frequency. Subjects rapidly modified their formant values over the course of fewer than ten trials to reach their maximum compensation value. However, on average the maximum adjustment was only a partial compensation, amounting to less than 30% of the perturbation magnitude.

The results support the view (Pick *et al.* 1989) that the auditory concomitant of speech production is used in a control system to feed back and modify vowel production. This auditory feedback is part of the standard operational principles underlying vowel production. The present results extend this conclusion to the control of formant frequency. Our findings are consistent with data from perturbations to other facets of acoustic speech feedback (Burnett *et al.*, 1998; Hain *et al.*, 2000; Pick *et al.*, 1989) as well as with data from the study of other overlearned motor behaviors [e.g., postural responses; see Weerdesteyn *et al.*, (2008)]. Compensatory behavior is not easily eliminated by instructions.

In the study of voluntary control of the Lombard effect (Pick *et al.*, 1989), it was shown that subjects were not spontaneously able to inhibit increasing their vocal amplitude as background noise level increased, but they could be trained to do so if provided with sufficient feedback and strategies. However, even under these special conditions, the subjects appear to have gained this control using a strategy of generally reducing vocal amplitude in all background noise levels. Skilled singers have experienced extensive training of voice level and pitch control. Even with this extensive training, although the magnitude of compensatory responses to vocal frequency perturbations is reduced, compensation is not completely eliminated (Zarate and Zatorre, 2008). None of the talkers in the current study had experienced formant-shifted feedback before participating in the experiment. Thus, it is possible that through training, the magnitude of compensation might be reduced.

The compensation to altered auditory feedback observed in the present experiment might serve many useful functions. Ongoing learning might be needed to stabilize a representation of the speech motor system that is used in a predictive fashion to control rapid movements. This is consistent with proposals for the role of "internal models" as part of motor planning and control (Wolpert and Kawato, 1998). Plasticity is not a necessary feature of this proposal, but some type of corrective mechanism is. Another possible function is necessitated by the fact that the vocal tract changes in morphology continuously over the lifespan (Fitch and Giedd, 1999; Vorperian *et al.*, 2005). While the major morphological changes happen before the age of 20, many structures and cavities continue to change later in life, and therefore we must possess some mechanism that is able to modify articulatory goals. Such adaptation may be part of a more general plasticity in motor learning observed for even the simplest sensorimotor behaviors (Wolpaw, 2007) and which supports the learning of new skilled activity and recovery from injury. Finally, these auditory-vocal adjustments may serve to tune speech production to ongoing changes in background acoustic conditions to ensure intelligibility.

The functions and mechanisms of such plasticity in speech production are unclear in part because it is still not known at what level of the speech motor system such changes occur. In a review of visual-motor perturbation studies that spanned more than 50 years, Epstein (1967) identified six possible alternative sites of adaptation in those studies. Of these alternatives, five are readily adaptable to the auditory feedback context. First, the adaptive changes could

be strictly auditory. Speech sound categories have been shown to be modifiable in a variety of learning contexts [e.g., see Kraljic *et al.* (2008), Maye *et al.* (2008), Norris *et al.* (2003), van Linden and Vroomen (2007), and Davis *et al.* (2005)]. The diversity of conditions that produce such changes in speech perception (lexical status, visual speech, acoustic experience, pragmatic information, etc.) indicates an auditory speech perception system that is inherently dynamic. The relationship of this system to the auditory processes supporting speech production is less well understood with only a few demonstrations of an association between the two types of perception [e.g., see Cooper *et al.* (1976) and Newman (2003)]. Second, the adaptive changes could be strictly proprioceptive. Ostry and co-workers (Tremblay *et al.*, 2003; Nassir and Ostry, 2006) showed that jaw movements during speech adapt to changes in the dynamic force field. This learning can occur without the presence of acoustic changes in the speech and presumably involves a representation of speech movement at the proprioceptive level. Cutaneous, joint, and muscle receptors provide a rich sensory representation of articulator states. Third, the adaptive changes could be strictly motoric. The tongue muscles act synergistically during vowel production to control the shape and position of the vocal tract constriction (Perkell, 1996), and any changes in resonance properties caused by the feedback perturbation would necessitate a modification of this organization. Fourth, while these specific sensory and motor changes are possible, perhaps the most likely account is that the learning is of a more multimodal nature. One multimodal change could involve sensory-motor "recorrelation," as has been suggested for prism adaptation studies [e.g., see Epstein (1967)] and some auditory-motor speech learning studies [e.g., see Purcell and Munhall (2006)]. For the study under discussion here, this would mean learning a new mapping between vocal tract shape and speech acoustics during the perturbation phase of the experiment and then relearning the old mapping again during the return phase.

The fifth and final alternative is that the observed changes could involve "conscious correction." The present study addressed this option, and the results suggest that this possibility is unlikely; instead, formant compensation is robust under a variety of different instructional sets. We do not think that our data are evidence of a fixed response system that cannot be changed with practice or strategies. The tendency to describe compensatory behavior of the kind observed here as reflexive or automatic must be tempered by the growing recognition that even the most widely accepted examples of "reflexes" such as the tendon jerk can be modified or conditioned (Wolpaw, 2007). Rather, the present results are evidence that compensatory responses to vowel perturbation are not simply overt strategic responses to detecting manipulated feedback. The presence of aftereffects that persist beyond the perturbation is itself a strong argument that strategic compensation is an unlikely explanation for these data (Epstein, 1967). These aftereffects can be seen in Fig. 2. The formant values do not immediately return to baseline levels when normal feedback is provided.

Inference of mechanism from behavioral data is limited since the same observed behavior could be accomplished in different ways. For example, the same behavior could be accomplished by relying on different neural substrates or on the same substrate but in varying amounts. Zarate and Zatorre (2008) reported that when singers and nonsingers were instructed to compensate for pitch perturbations, both performed this task equally well. However, neuroimaging revealed that singers exhibited more activity in the anterior cingulate cortex, superior temporal sulcus, and putamen than did nonsingers. In a study of the stereotypical balance-recovery response in humans, Weerdesteyn *et al.* (2008) found that subjects could voluntarily inhibit a stepping response for balance recovery when instructed to do so. However, electromyographic records showed that both the balance-recovery trials and the trials in which the subjects were instructed to inhibit the response and fall forward had similar compensatory muscle activation patterns, with similar latencies but dramatically different activation amplitudes. Weerdersteyn *et al.* (2008) suggested that a consistent balance-recovery response is always generated but that the magnitude of the response can be regulated for different goals. Thus, a more complete understanding of the mechanisms supporting the compensatory speech response will require a more detailed physiological investigation.

In summary, the data presented here indicate that motor planning and control of vowel production must incorporate the auditory consequences of the movements as feedback. Modifications of this acoustic signal that result in an error relative to expected sensory feedback initiate compensatory behavior even when subjects are aware of the manipulation. That such compensation appears obligatory suggests that, in everyday life, compensation happens automatically in response to changing acoustic conditions contributing to optimal intelligibility even while the talker is unaware of the process.

ACKNOWLEDGMENTS

Bauer, J. J., Mittal, J., Larson, C. R., and Hain, T. C. (**2006**). "Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude," J. Acoust. Soc. Am. **119**, 2363–2371.

Brainard, M. S., and Doupe, A. J. (**2000**). "Auditory feedback in learning and maintenance of vocal behaviour," Nat. Rev. Neurosci. **1**, 31–40.

Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (**1998**). "Voice F0 responses to manipulations in pitch feedback," J. Acoust. Soc. Am. **103**, 3153–3161.

Cooper, W. E., Billings, D. and Cole, R. A. (**1976**). "Articulatory effects on speech perception: A second report," J. Phonetics **4**, 219–232.

Cowie, R., and Douglas-Cowie, E. (**1992**). *Postlingually Acquired Deafness: Speech Deterioration and the Wider Consequences* (Mouton de Gruyter, New York).

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., and McGettigan, C. (**2005**). "Learning to understand noise-vocoded speech. Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences," J. Exp. Psychol. Gen. **134**, 222–241.

Epstein, W. (**1967**). *Varieties of Perceptual Learning* (McGraw-Hill, New York).

Fitch, W. T., and Giedd, J. (**1999**). "Morphology and development of the

human vocal tract: A study using magnetic resonance imaging," J. Acoust. Soc. Am. **106**(3), 1511–1522.

Gracco, V. L. and Abbs, J. H. (**1985**). "Dynamic control of the perioral system during speech: Kinematic analyses of autogenic and nonautogenic sensorimotor processes," J. Neurophysiol. **54**, 418–432.

Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., and Kenney, M. K. (**2000**). "Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex," Exp. Brain Res. **130**, 133–41.

Houde, J. F., and Jordan, M. I. (**1998**). "Sensorimotor adaptation in speech production," Science **279**, 1213–1216.

Kawahara, H. (**1995**). "Transformed auditory feedback: The collection of data from 1993.1 to 1994.12 by a new set of analysis procedures," TRH-120, ATR Human Information Processing Research Laboratories, Kyoto, pp. 1–52.

Kraljic, T., Samuel, A. G., and Brennan, S. E. (**2008**). "First impressions and last resorts: How listeners adjust to speaker variability," Psychol. Sci. **19**, 332-338.

Lane, H., and Tranel, B. (**1971**). "Lombard Sign and Role of Hearing in Speech," J. Speech Hear. Res. **14**, 677–709.

Maye, J., Weiss, D. J., and Aslin, R. N. (**2008**). "Statistical phonetic learning in infants: Facilitation and feature generalization," Dev. Sci. **11**, 122–134.

Nassir, S. M., and Ostry, D. J. (**2006**). "Somatosensory precision in speech production," Curr. Biol. **16**, 1918–1923.

Newman, R. (**2003**). "Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report," J. Acoust. Soc. Am. **113**, 2850–2860.

Norris, D., McQueen, J. M., and Cutler, A. (**2003**) "Perceptual learning in speech," Cogn. Psychol. **47**, 204–238.

Orfandidis, S. J. (**1988**). *Optimum Signal Processing: An Introduction* (Mac-Millan, New York).

Perkell, J. S. (**1996**). "Properties of the tongue help to define vowel categories: Hypotheses based on physiologically-oriented modeling," J. Phonetics **24**, 3–22.

Pick, H. L., Jr., Siegel, G. M., Fox, P. W., Garber, S. R., and Kearney, J. K. (**1989**). "Inhibiting the Lombard effect," J. Acoust. Soc. Am. **85**, 894–900.

Prochazka, A., Clarac, F., Loeb, G. E., Rothwell, J. C., and Wolpaw, J. R.. (**2000**). "What do reflex and voluntary mean? Modern views on an ancient debate," Exp. Brain Res. **130**, 417–432.

Purcell, D. W., and Munhall, K. G. (**2006**). "Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation," J. Acoust. Soc. Am. **120**, 966–977.

Scheifele, P. M., Andrew, S., Cooper, R. A., Darre, M., Musiek, F. E., and Max, L. (**2005**). "Indication of a Lombard vocal response in the St. Lawrence River beluga," J. Acoust. Soc. Am. **117**, 1486–1492.

Smotherman, M. (**2007**). "Sensory feedback control of mammalian vocalizations," Behav. Brain Res. **182**, 315–326.

Tremblay, S., Shiller, D. M. and Ostry, D. J. (**2003**). "Somatosensory basis of speech production," Nature (London) **423**, 866–869.

Van Linden, S., and Vroomen, J. (**2007**). "Recalibration of phonetic categories by lipread speech versus lexical information," J. Exp. Psychol. Hum. Percept. Perform. **33**, 1483–1494.

Villacorta, V. M., Perkell, J. S., and Guenther, F. H. (**2007**). "Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception," J. Acoust. Soc. Am. **122**, 2306–2319.

Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., and Yandell, B. S. (**2005**). "Development of vocal tract length during early childhood: A magnetic resonance imaging study," J. Acoust. Soc. Am. **117**, 338–350.

Weerdesteyn, V., Laing, A. C., and Robinovitch, S. N. (**2008**). "Automated postural responses are modified in a functional manner by instruction," Exp. Brain Res. **186**, 571–580.

Wolpaw, J. R. (**2007**). "Spinal cord plasticity in acquisition and maintenance of motor skills," Acta Physiol. **189**, 155–169.

Wolpert, D. M., and Kawato, M. (**1998**). "Multiple paired forward and inverse models for motor control," Neural Networks **11**, 1317–1329.

Zarate, J. M., and Zatorre, R. J. (**2008**). "Experience-dependent neural substrates involved in vocal pitch regulation during singing," Neuroimage **40**, 1871–1887.

Munhall *et al.*: Instructions and response to formant perturbation