Probing the independence of formant control using altered auditory feedback

Ewen N. MacDonald[a)]

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

David W. Purcell

School of Communication Sciences and Disorders, University of Western Ontario, 1201 Western Road, London, Ontario, N6G 1H1, Canada

Kevin G. Munhall[b)]

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

Two auditory feedback perturbation experiments were conducted to examine the nature of control of the first two formants in vowels. In the first experiment, talkers heard their auditory feedback with either F1 or F2 shifted in frequency. Talkers altered production of the perturbed formant by changing its frequency in the opposite direction to the perturbation but did not produce a correlated alteration of the unperturbed formant. Thus, the motor control system is capable of fine-grained independent control of F1 and F2. In the second experiment, a large meta-analysis was conducted on data from talkers who received feedback where both F1 and F2 had been perturbed. A moderate correlation was found between individual compensations in F1 and F2 suggesting that the control of F1 and F2 is processed in a common manner at some level. While a wide range of individual compensation magnitudes were observed, no significant correlations were found between individuals' compensations and vowel space differences. Similarly, no significant correlations were found between individuals' compensations and variability in normal vowel production. Further, when receiving normal auditory feedback, most of the population exhibited no significant correlation between the natural variation in production of F1 and F2. © *2011 Acoustical Society of America.* [DOI: 10.1121/1.3531932]

I. INTRODUCTION

When we watch people's movements, it is natural to think of them as indivisible wholes   each movement the enactment of a single intention. However, movements have underlying structure with dimensions or independent degrees of freedom that are an essential aspect of planning and control. For example, reaching movements are thought by some (e.g., Ghez *et al.*, 1997) to be planned as vectors with direction and extent of movement of the hand as separate control dimensions. In this paper we explore the independence of the control of individual formants during vowel production using an auditory feedback paradigm. Specifically, we test whether the feedback control of speech involves independent corrections for perturbations to individual formants. By doing so, we test the specificity or granularity of speech motor control.

From standard descriptions in either articulatory or auditory terms, vowels can be well characterized by a small num-ber of parameters. Since the earliest studies of spectrography of speech (e.g., Joos, 1948) vowels have been classified in a space defined by the first two formants of the vowel (F1, F2). The axes of this space correspond to the traditional phonetic dimensions of vowel height and front-backness, though the mapping to the actual tongue height position is not straightforward (Ladefoged, 1982; Wood, 1982; Ladefoged and Maddieson, 1996). A separate line of work on tongue configurations has sought to understand how many factors are needed to account for the tongue's shape during vowel production (Harshman *et al.*, 1977; Jackson, 1988; Maeda, 1990; Nix *et al.*, 1996; Hoole, 1998; Beautemps *et al.*, 2001). These studies have consistently shown that only a small number of components are required to account for the variance in tongue shapes within and across languages.

The idea that a small number of dimensions of articulation are involved in the production of individual speech sounds reflects the belief that the nervous systems strives to control the vocal tract using muscle synergies (e.g., Fowler and Saltzman, 1993). Synergies are groupings or couplings of muscles that have consistent spatiotemporal activation patterns (e.g., Ting and McKay, 2007). For both speech and non-speech movements, synergies have long been proposed as a solution to the degrees of freedom problem (Bernstein, 1967). The nervous system is thought to reduce the

---

a)Author to whom correspondence should be addressed. Also at: Department of Electrical and Computer Engineering, Queen's University, Humphrey Hall, 62 Arch Street, Kingston, Ontario, K7L 3N6, Canada. Electronic mail: ewen.macdonald@queensu.ca

b)Also at: Department of Otolaryngology, Queen's University, Humphrey Hall, 62 Arch Street, Kingston, Ontario, K7L 3N6, Canada.

computational task in movement by controlling groups of muscles as units instead of individually controlling the hundreds of motor units involved in a behavior. Thus, in speech, constrictions that are important for the production of vowel sounds can be defined as tasks shared across coupled articulators and muscles (Saltzman and Munhall, 1989).

In spite of the long history investigating synergies in the study of motor control, there are still many unknowns about how such couplings could work in coordination (Ting, 2007), or even whether synergies are the best way to characterize motor organization (Tresch and Jarc, 2009). One approach that has been important in this controversy has been the examination of both the immediate response to unexpected perturbations to the motor system and the degree to which the task is achieved in the response. Mechanical perturbations, for example, have been applied to the lips and jaw and compensatory responses have been observed in the same or biomechanically linked structures (Abbs and Gracco, 1984; Kelso et al., 1984; Gracco and Abbs, 1985; Shaiman, 1989) as well as more remote articulators such as the larynx (Munhall et al., 1994).

In this study we will use perturbations of auditory feedback to examine the relationship between the first two formants in vowels. Our aim is not to address the controversy about whether gestural or acoustic primitives are the best way to characterize speech goals (Perkell et al., 2000; Diehl et al., 2004; Browman and Goldstein, 1986; Fowler and Saltzman, 1993) but rather to use sensitivity to auditory feedback changes as a probe of the nature of the vowel control system. By examining the changes in response to formant perturbations, we can assess if vowel articulation exhibits synergistic control. A study in which auditory feedback is perturbed and where articulatory measures are also taken (e.g., Nieto-Castanon et al., 2005) would be required to better characterize speech goals.

Auditory feedback has been shown to be an important part of the sensorimotor control of speech production. Clinical evidence has shown that speech development is impaired in children with profound hearing loss (Oller and Eilers, 1988) and that postlingually deafened adults show deterioration of articulation (Cowie and Douglas-Cowie, 1992). In addition, a range of experimental techniques that involve real-time perturbations of auditory feedback show rapid modifications in speech production. When talkers receive delayed auditory feedback, they slow their speech and may become disfluent (Lee, 1950; Yates, 1963; Howell and Archer, 1984). When talkers receive feedback in which the loudness of their voice (Bauer et al., 2006) or background sound level is increased (Lombard, 1911; Lane and Tranel, 1971; Pick et al., 1989), talkers modify the loudness of their speech. When the fundamental frequency (Kawahara, 1995; Burnett et al., 1988; Jones and Munhall, 2000) or the vowel formants (Houde and Jordan, 2002; Purcell and Munhall, 2006; Villacorta et al., 2007; Munhall et al., 2009) are shifted up or down in frequency, subjects compensate by changing the corresponding acoustic parameters in the direction opposite to the perturbation. These frequency compensations persist following return to normal feedback, indicating that some form of auditory-motor representation is involved in vowel production.

These representations are thought to include information about how coordinated muscles and articulators change the vocal tract as well as the acoustic and dynamic consequences of movements (Kawato, 1989; Guenther et al., 2006; Wolpert et al., 2001). By learning a detailed model of speech sound production, the nervous system can use this model to predict the consequences of actions and detect errors rapidly. The level of detail represented and the flexibility of compensatory strategies using such error correction processes are unknown.

At one extreme, the individual vowel "targets" may be represented as rigid synergies in which the muscles consistently change together to compensate for auditory error. On the other hand, the representation may entail a more graded and detailed model of the vocal tract and the effects of articulatory changes on the acoustics, which could be much like what are modeled with a nomogram (Lindblom and Sundberg, 1971). The latter type of representation would be useful if the nervous system was capable of specific articulatory responses to small acoustic errors.

In the present experiments, we look at the relationship between F1 and F2 under a variety of conditions. First, we manipulated the auditory feedback for F1 and F2 separately to examine the independence of control of the traditional height and front-backness dimensions. If selective changes in the auditory-motor mapping for a single formant can quickly be learned it suggests that the vowel representation is not a rigid synergy with coupled F1 and F2 values. Rather, the spatial adjustments required to compensate for a single formant perturbation would have to be realized separately. The second goal is to test whether the formants show related behavior when examined under similar perturbation conditions. In a large meta-analysis (experiment 2) we investigated whether the magnitude and direction of compensations are correlated for F1 and F2 when the auditory feedback for both formants is perturbed. Finally, we examined the trial-to-trial covariance of F1 and F2 in a large data set of baseline trials with normal auditory feedback to determine the degree to which the variance in the first two formant frequencies was related under natural production conditions. Data regarding this possibility have been reported in imitation studies (e.g., Vallabha and Tuller, 2004) which indicate independence of variation in F1 and F2 production.

These sets of analyses provide a framework for examining how vowel acoustics are controlled. By testing the potential independence (experiment 1) and the natural covariance of F1 and F2 (experiment 2), the present studies provide a clearer view of the planning and control of articulation.

II. GENERAL METHODS

A. Equipment

The equipment used in both experiments 1 and 2 was the same as that reported in Purcell and Munhall (2006) and Munhall et al. (2009). Participants were seated in a sound-insulated booth (Industrial Acoustic Company, Bronx, NY). Participants were instructed to say the words that appeared on a monitor in front of them at a natural rate and level. Each word prompt lasted 2.5 s and the inter-trial interval was approximately 1.5 s.

The speech was recorded using a headset microphone (Shure WH20), amplified (Tucker-Davis Technologies MA3

microphone amplifier), low-pass filtered with a cutoff frequency of 4500 Hz (Frequency Devices 901 filter), and digitized at 10 kHz (National Instruments PXI-8106 embedded controller). The National Instruments system generated a new formant estimate every nine speech samples. Using this estimate, filter coefficients were calculated to produce formant shifts. To mask bone-conducted feedback, the manipulated voice signal was amplified and mixed with speech noise (Madsen Midimate 622 audiometer) and presented over headphones (Sennheiser HD 265) such that the speech and noise were presented at approximately 80 and 50 dBA sound pressure level (SPL), respectively.

B. Online formant shifting and detection of voicing

Detection of voicing and formant shifting was performed as previously described in Munhall *et al*. (2009). Voicing was detected using a statistical amplitude-threshold technique. The formant shifting was achieved in real-time using an infinite impulse response (IIR) filter. Formants were estimated every 900 μs using an iterative Burg algorithm (Orfanidis, 1988). Filter coefficients were computed based on these estimates such that a pair of spectral zeroes was placed at the location of the existing formant frequency and a pair of spectral poles was placed at the desired frequency of the new formant.

C. Estimating model order

The iterative Burg algorithm used to estimate formant frequencies requires a parameter, the model order, to determine the number of coefficients used in the auto-regressive (AR) analysis. Prior to data collection, talkers ran through a vowel screener procedure in which they produced six utterances of seven English vowels in an /hVd/ context ("heed," "hid," "hayed," "head," "had," "hawed," and "who'd"). These utterances were analyzed with model orders ranging from 8 to 12. The best model order was selected using a heuristic based on minimum variance in formant frequency over a 25 ms segment midway through the vowel.

D. Offline formant analysis

The recorded data were analyzed in the same way as that used by Munhall *et al*. (2009). The boundaries of the vowel segment in each utterance were estimated using an automated process based on the harmonicity of the power spectrum. These boundaries were then inspected by hand and corrected if required.

The first three formant frequencies were estimated offline from the first 25 ms of a vowel segment using a similar algorithm to that used in online shifting. The formants were estimated again after shifting the window 1 ms and repeated until the end of the vowel segment was reached. For each vowel segment, a single "steady-state" value for each formant was calculated by averaging the estimates for that formant from 40% to 80% of the way through the vowel. While using the best model order reduced gross errors in tracking, occasionally one of the formants was incorrectly categorized as another (e.g., F2 being misinterpreted as F1, etc.). These incorrectly categorized estimates were found and corrected

by examining a plot with all the steady-state F1, F2, and F3 estimates over time for each individual.

III. EXPERIMENT 1

When subjects are given the task of imitating tokens from a vowel continuum, they show systematic errors, or biases, in their production of F1 and F2 (e.g., Repp and Williams, 1987; Vallabha and Tuller, 2004). The source of these vowel production tendencies may lie in both perception and production but in either event this pattern indicates possible constraints on the ways F1 and F2 can covary for a talker. In this experiment, we address this by unexpectedly modifying the auditory feedback for the first two formants separately. Four groups of subjects produced the monosyllabic English word "head" under conditions of normal feedback and when one of the formants was shifted up or down in frequency. For each group of subjects only one formant was shifted in one direction but by comparing the behavior of both F1 and F2 across the four groups, we can assess the degree to which the feedback system is capable of producing local changes in the vowel spectrum. Figure 1 shows the average within-subject variance in F1 and F2 production along with the mean locations for normal productions of the vowels /ɪ/, /ɛ/, and /æ/ from all four groups of subjects. The four perturbation vectors are superimposed on this F1 and F2 space. Note that the individual formant perturbations project the feedback for the vowel /ɛ/ outside of the "path" connecting the vowels in this region of the vowel space. English front vowels covary in F1 and F2 and their average values for vowels produce an angular path in F1/F2 space. Lower vowels like /æ/ are more back than high vowels like /i/. If the compensation trajectories can freely deviate from the vowel space path equally for all directions of perturbation, it will suggest that the control space for processing auditory feedback is not constrained by
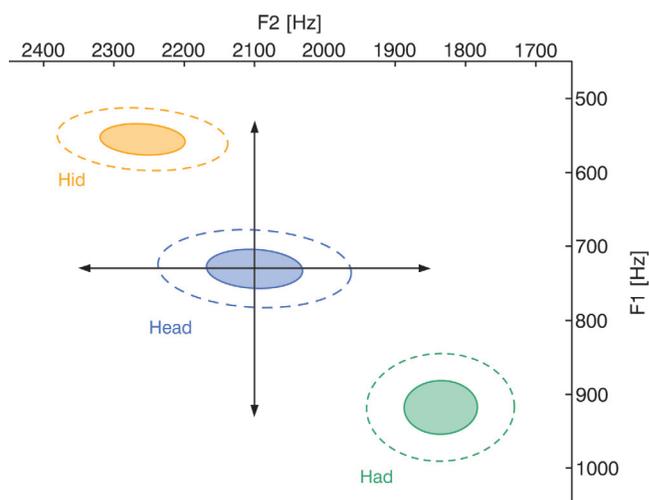

FIG. 1. (Color online) Schematic of perturbations used in experiment 1 in a vowel space context. Each pair of concentric ellipses indicates the distribution of an average talker's production of /ɪ/, /ɛ/, and /æ/ in an /hVd/ context. The center of each pair of ellipses indicates the mean production of an average individual and the solid and dashed ellipses indicate one and two standard deviations, respectively. The four arrows indicate the auditory feedback perturbations the four groups received.

the location of vowel categories and the covariance of F1 and F2 for those categories.

A. Participants

The participants in experiment 1 consisted of 48 female undergraduate students (mean age of 18.9 yr, range: 18 24 yr.). All talkers spoke English as their first language, reported no speech or language impairments, and were screened to ensure audiometric thresholds were normal [<25 dB hearing level (HL) over a range of 500 4000 Hz]. The protocol for this study was approved by the institutional ethics review board and all talkers provided informed consent.

B. Procedure

The talkers were randomly assigned to one of four groups (F1+200, F1−200, F2+250, or F2−250. During the experiment, all four groups produced the English word "head" a total of 120 times. The experiment consisted of three phases. In the first phase, Baseline, 20 utterances were produced with normal feedback (i.e., amplified but no shift in formant frequency) to estimate baseline F1 and F2 values. In the second phase, Shift, 50 utterances were produced with feedback in which either the frequency of F1 or F2 was shifted. For the F1+200 and F1−200 groups, the F1 frequencies were increased or decreased by 200 Hz, respectively. For the F2+250 and F2−250 groups, the F2 frequencies were increased or decreased by 250 Hz, respectively. In the final phase, Return, 50 utterances were produced with normal feedback (i.e., the formant shift was abruptly turned off).

C. Results and discussion

Estimates of average baseline production formant frequencies were calculated for each individual based on the mean of the last 15 utterances of the Baseline phase (i.e., utterances 6 20). The first five utterances were eliminated in an attempt to minimize effects of initial familiarization to the task and headphones. The formant estimates were then normalized for each individual by subtracting the talker's baseline average. The normalized results for each utterance, averaged across the talkers in each group, can be seen in Fig. 2.

The four groups of talkers all showed compensations in the perturbed formant. During the Shift phase each group altered the frequency of the formant with perturbed feedback in a direction opposite that of the perturbation. However, for the formant that was not perturbed, F1 and F2 showed different patterns.

In response to the perturbation in F1 during the Shift phase, both groups (F1+200 and F1−200) altered production of F1 by approximately 50 Hz in a direction opposite that of the perturbation. As well, both groups altered production of the unperturbed formant, F2, decreasing average production by approximately 30 and 40 Hz for the F1+200 and F1−200 groups, respectively. However, this decrease in F2 production was more gradual than the abrupt change observed when F2 was perturbed or the similarly abrupt change observed in F1 when F1 was perturbed. For both groups, when feedback was returned to normal at the beginning of the Return phase, the
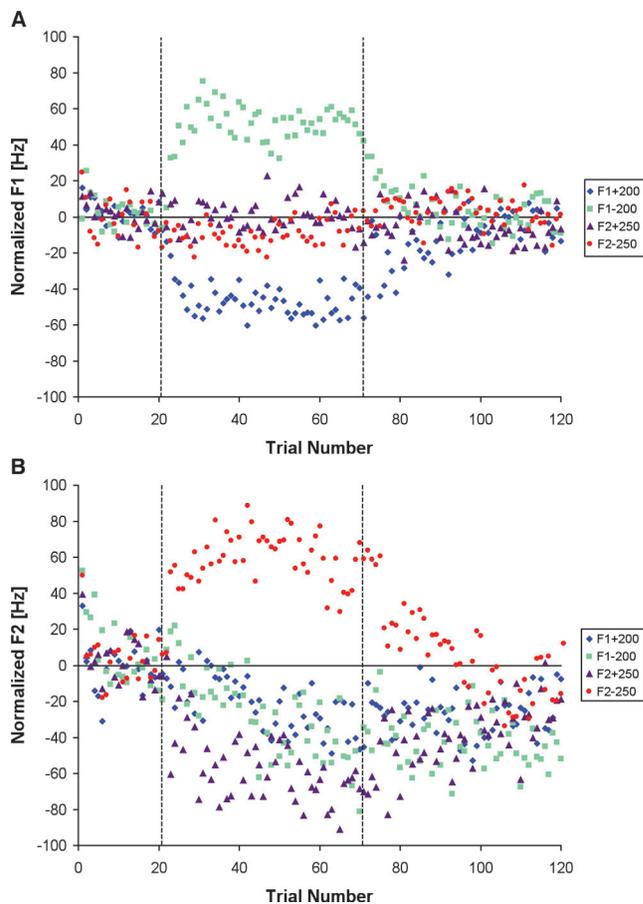

FIG. 2. (Color online) Average normalized formant frequency for (a) F1 and (b) F2 over the course of the experiment. The two vertical dashed lines indicate when the perturbed feedback was introduced and when auditory feedback was returned to normal.

production of F1 returned to baseline but the production of F2 remained the same as it was at the end of the Shift phase. This second observation indicates that the motor control system can control F2 independently of F1. When taken together, these results suggest that the change in F2 is of a different character than the F1 response to the perturbation.

In response to the perturbation of F2 during the Shift phase, both groups (F2+250 and F2−250) altered production of F2 by approximately 65 Hz in a direction opposite that of the perturbation. However, unlike the results of the other two groups, production of the unperturbed formant, F1, remained unchanged across the three phases of the experiment. When the perturbation was removed, production in F2 returned toward the baseline average. However, at the end of the Return phase, production of F2 for both groups was below the Baseline average.

To quantify the change in production and statistically test the observations reported above, three intervals were defined based on the last 15 utterances of each phase (i.e., utterances 6 20 for Baseline, 56 70 for Shift, and 106 120 for Return). In all intervals, it is assumed that formant production has reached steady state for the respective phase. The non-normalized F1 and F2 estimates in each interval were averaged for each individual and used in the analyses. Repeated measure analyses of variance (ANOVAs) were

MacDonald *et al.*: Independence of formant control

conducted separately for the F1 and F2 results, with phase as a within-subjects and group as a between-subjects factor.

For the F1 results, a significant main effect was found for group [$F(3,44) = 7.685$, $p < 0.001$] but not for phase [$F(2,88) = 0.199$, $p = 0.82$]. This statistical pattern was observed because the responses of the F1+200 and F1−200 groups to the F1 perturbation were in opposite directions because the perturbations were in opposite directions in frequency. A significant interaction of group × phase was found [$F(6,88) = 14.302$, $p < 0.001$]. In the context of this experiment, a type II error is of more concern that a type I error as many previous experiments have demonstrated that talkers alter production of a perturbed formant (e.g., Houde and Jordan, 2002; Purcell and Munhall, 2006; Villacorta et al., 2007; Munhall et al., 2009; MacDonald et al., 2010). Thus, multiple comparisons with no adjustment were conducted for the F1 results. Significant differences between the Baseline and Shift phases and between the Shift and Return phases were found for both the conditions in which F1 was perturbed ($p < 0.05$ for both F1+200 and F1−200 groups) but not between the two phases for the conditions where F1 was not perturbed ($p > 0.05$ for both F2+250 or F2−250 groups).

The results in F2 were slightly different. Using Greenhouse-Geisser correction, a significant main effect was found for both group [$F(3,44) = 2.855$, $p < 0.05$] and phase [$F(1.585,88) = 12.694$, $p < 0.001$]. Further, interaction of group × phase was also significant [$F(4.755,88) = 12.818$, $p < 0.001$]. As with the F1 results, multiple comparisons with no correction were conducted for the F2 results. For the two conditions where F2 was perturbed, significant differences were found between the Baseline and Shift phases and also the Shift and Return phases ($p < 0.05$ for both F2+250 and F2−250 groups). For the two conditions where F1 (but not F2) was perturbed, the differences between the Baseline and Shift phases were significant ($p < 0.05$ for both F1+200 and for F1−200) though the differences between the Shift and Return phases were not ($p > 0.05$ for both F1+200 and F1−200). Thus, the two groups that received perturbed F1 feedback altered production of F2 during the shift phase, but F2 production remained constant when the F1 perturbation was removed.

Previous studies have demonstrated that talkers alter production of both F1 and F2 in response to altered auditory feedback in which both formant frequencies have been perturbed (Houde and Jordan, 2002; Munhall et al., 2009; MacDonald et al., 2010). Thus, perturbing the auditory feedback of formants induces talkers to alter production of those formants. However, when a single formant was changed, the compensatory behavior showed some specificity. Villacorta et al. (2007) gradually shifted F1 and examined production of both F1 and F2. While their results showed talkers compensated for the perturbation in F1, no change in the production of F2 was observed. In the present study, no change in the production of F1 was observed when F2 was abruptly perturbed but a change in F2 was observed when a sudden perturbation to F1 was applied. However, the change in F2 that we observed appears to be of a different character than the response in F1 and thus due to some other, as yet unexplained, reason. We reach this conclusion based on the con-

sistent direction of F2 change independent of the direction of F1 perturbation, the more gradual change in F2 compared to that in F1, and the failure of F2 to change when the feedback perturbation was removed. Given the different directions of compensation in F1 for the two F1 perturbations, any indirect effect on F2 by changes in jaw height or tongue constriction, for example, would have been sensitive to the direction of compensation (e.g., Lindblom and Sundberg, 1971). Thus we conclude that the change in F2 in response to the F1 perturbation does not suggest that the control is dependent. Regardless, it is uncertain why the change in F2 observed here when the perturbation of F1 was applied was not observed by Villacorta et al.

The data as a whole suggest that auditory feedback changes can induce quite specific modifications in articulation. In this case, a change of the vocal tract shape to bring about a shift in only one of the formants was made. The question of how specific and accurate speech motor control is has been the focus of a variety of studies. Ideas such as the quantal characteristics of the vocal tract and saturation effects of motor output can lead to the conclusion that perceptual stability can be produced without the need for excessive movement precision (Perkell et al., 2000). However, detailed biomechanical modeling of the human tongue shows that variations in individual muscle activity produce significant influences on the formants even in the context of potential saturation effects (Buchaillard et al., 2009). Buchaillard et al. conclude that this variability must be actively reduced for perceptual efficacy (see also Mooshammer et al., 2001). This kind of effect suggests an active precision that is far greater than what the small number of degrees of freedom posited would suggest (e.g., Beautemps et al., 2001).

From the point of view of the questions addressed by Buchaillard et al. (2009), the present test is crude. However our results do show that feedback errors in a localized region of the spectrum can be detected and independently accommodated. This does not mean that the standard operation of speech motor planning and execution exhibit the same degree of independence for the movements that produce the first and second formants in the altered feedback context tested here. Feedback error correction may have unique status in coordination.

IV. EXPERIMENT 2

The results of experiment 1 demonstrated that talkers are, on average, able to compensate for perturbations in feedback of F1 and F2 independently. When faced with a local perturbation to a single formant, a talker could produce compensations that did not require a global change to vowel articulation. An alternative way to look at the question of the specificity of control in speech is to measure the way in which the formants naturally covary. In this experiment, we examine covariation of the formants in two contexts. First we look at the degree to which the compensatory behavior of the first two formants is correlated when both are perturbed. Previous studies have demonstrated large individual differences in the magnitude of compensation in both F1 and F2 (Munhall et al., 2009). Since perturbations evoke only partial compensations, differences in the size of the response can be

J. Acoust. Soc. Am., Vol. 129, No. 2, February 2011

MacDonald et al.: Independence of formant control   959

viewed as differences in the feedback gain employed by the control system leading to different steady-state error. In the context of the current studies, we examine if talkers use the same gain parameter with each formant by conducting a meta-analysis on a large set of previously collected data. Thus, while experiment 1 demonstrated that the speech motor control system is capable, under some conditions, of independent control of F1 and F2, experiment 2 investigates if the control system processes F1 and F2 errors in an identical manner when experimental feedback manipulations are used. In addition we use this large meta-analysis to examine the natural variability in production of non-perturbed utterances. Even the most skilled and practiced motor behaviors show variability that makes voluntary movements not completely repeatable and this imposes limits on movement accuracy. By examining the covariance structure of the same utterance in unperturbed conditions we can test for constraints in motor planning and execution (van Beers, 2009). Finally, we also use this large meta-analysis to explore possible origins of between talker variability in the magnitude of compensation. Correlations between two variables and the magnitude of compensation were conducted: (1) between-vowel average distance and (2) standard deviation in normal production. Together these analyses may reveal the natural operating principles of formant production.

A. Experimental conditions

The data used in the meta-analysis derive from seven experiments that had been run previously in our lab. The data from these conditions had been collected to investigate the specificity of compensation in response to real-time formant shifting of the vowel /ɛ/. While the specific details varied across experiment, there are some important features common to all. In three of the experiments, talkers alternated saying "head" and "hid" (an utterance of one word was followed by an utterance of the other). In the other four experiments, after each utterance of the word "head," talkers would hear a pre-recorded utterance of either the word "hid" or "head." All of the experiments included a baseline phase in which talkers would produce at least 20 utterances of "head" with normal feedback and a shift phase with at least 30 utterances of "head" with altered feedback in which F1 was increased by 200 Hz and F2 was decreased by 250 Hz.

A total of 116 female undergraduate students (mean age of 19.0 yr, range: 17 24 yr.) participated in seven experiments. All talkers spoke English as their first language, reported no speech or language impairments, and were screened to ensure audiometric thresholds were normal (<25 dB HL over a range of 500 4000 Hz). The protocol for these studies was approved by the institutional ethics review board and all talkers provided informed consent.

B. Results and discussion

In the experiments comprising the meta-analysis, all talkers received altered feedback in which F1 was increased by 200 Hz and F2 was decreased by 250 Hz. The effect of this perturbation was to shift the production of "head" to the average position of the word "had" in F1 F2 space. As seen in experiment 1, when talkers receive altered feedback in

which formants have been perturbed in real-time, talkers spontaneously alter production of the perturbed formant. In general, this compensatory behavior is incomplete with the magnitude of the change in production being smaller than that of the perturbation.

For each talker, compensation was computed for both F1 and F2 based on the difference between the average of the last 15 utterances of the baseline phase and the shift phase. The sign of the compensation was defined as positive if the change in production was opposed to that of the formant shift or negative if the compensation followed the direction of the shift.

Two univariate ANOVAs were conducted on the F1 and F2 compensation data with experiment as a factor. For both ANOVAs, no significant effect of experiment was found [$F(6,109) = 1.236$, $p = 0.29$ for F1 compensation; $F(6,109) = 1.045$, $p = 0.4$ for F2 compensation] indicating that the seven slightly different protocols did not produce different magnitudes of compensation. A histogram of individual compensations in F1 and F2 is plotted in Fig. 3.
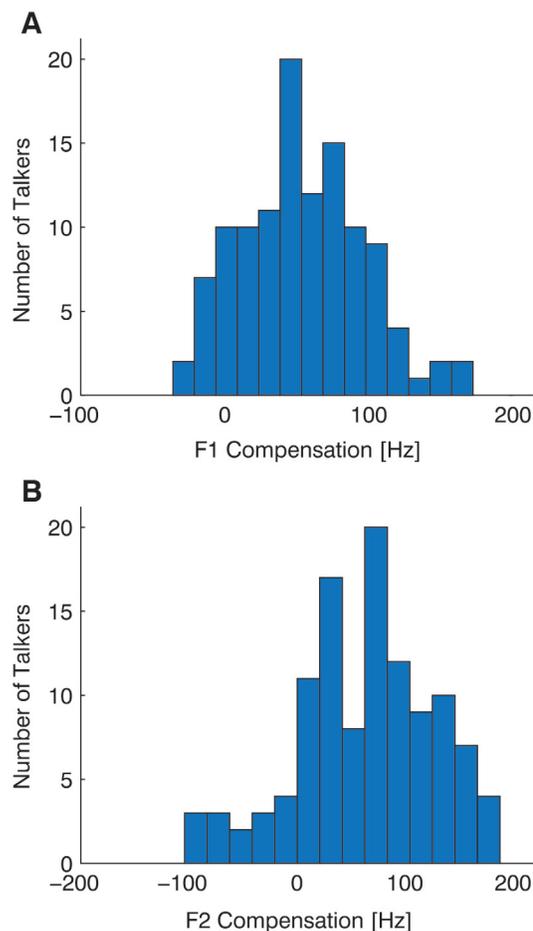

FIG. 3. (Color online) Histograms of individuals' compensation in (a) F1 and (b) F2. For each experimental condition, two intervals were defined. The base line interval was defined as the last 15 utterances spoken with normal feedback before a perturbation was introduced. The shift interval was defined as the last 15 utterances spoken with altered feedback. For each individual, the magni tude of the compensation was calculated from the difference in average form ant frequency between the two intervals. The sign of the compensation was defined as positive when the change in production is opposed to that of the formant shift and negative when it follows that of the formant shift.

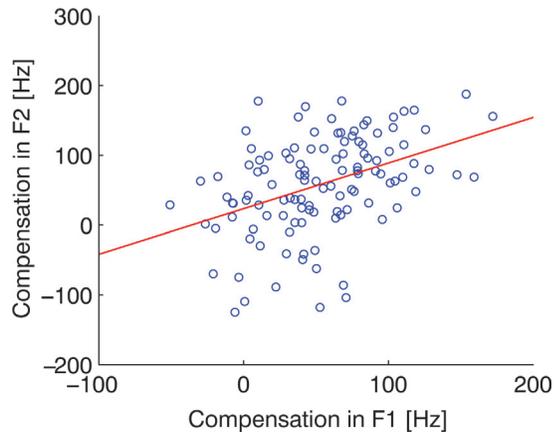MacDonald *et al.*: Independence of formant control

FIG. 4. (Color online) Scatter plot of individuals' compensations in F1 and F2. The solid line indicates the linear regression on the data.

For F1, the mean compensation was 53 Hz (26.5% of the perturbation) with a standard deviation of 44 Hz. For F2, the mean compensation was 58 Hz (23.2% of the perturbation) with a standard deviation of 69 Hz. For both compensation distributions, the standard deviation is similar in magnitude to the mean. From a behavioral perspective, this indicates a large variability across talkers. While the majority of the population exhibits compensation, there is a significant minority that exhibits little or no compensation. This is in accordance with previous studies that have described large inter-talker variability (Munhall et al., 2009).

To explore the relationship of the response to the perturbation, a scatter plot of the F1 and F2 compensations of each individual was produced and can be seen in Fig. 4. While individuals' compensation in F1 and F2 are correlated [$r(114) = 0.412$, $p = 0.001$], there is considerable variance around the regression line; only 16% of the variance in compensation in one formant is explained by compensation in the other. This indicates only a weak coupling between compensatory behaviors across formants.

When exposed to altered auditory feedback in which formants have been perturbed, not all talkers exhibit compensatory behavior. Previous studies have observed that some talkers alter production in a manner that follows (rather than opposes) the direction of the perturbation in pitch (Burnett et al., 1998) or in formant frequency (Munhall et al., 2009; MacDonald et al., 2010). From the meta-analysis, 26 (22.4%) of the 116 talkers exhibited following behavior with 8 (7%) following in F1 alone, 14 (12.1%) in F2 alone, and 4 (3.4%) in both F1 and F2. Thus, a minority of the talkers (22 talkers, 19.0%) compensated in one formant but followed in the other.

From vowel space data collected during the experiments, we can examine the possibility that the degree of compensation in one formant is influenced by individual differences in the vowel space. Prior to carrying out the experiments, talkers were run through a vowel screener procedure to estimate the model order used by the real-time formant shifter. In this procedure, utterances of other vowels in a /hVd/ context were collected. Using these data, the average F1 and F2 of the seven vowels were calculated for each individual (see Table I). A correlation analysis was conducted between an individual's compensation in one formant and that individual's formant difference between /ɛ/ and the two neighboring vowels /ɪ/ and /æ/. For both F1 and F2, no significant correlations were found between compensation and the difference in average productions of /ɛ/ and /æ/. Similarly, no significant correlations were found between compensation and the difference in average production of /ɪ/ and /ɛ/. These results are similar to findings in previous studies (MacDonald et al., 2010) and suggest that individual differences in the spacing between adjacent vowels does not affect compensation.

Previous studies have investigated the relationship between speech perception and production and found that talkers with lower formant discrimination thresholds have lower variance in vowel production (Perkell et al., 2004) and larger compensation in response to altered formant feedback (Villacorta et al., 2007). In relating those results to the directions into velocities of articulators (DIVA) model of speech motor control (Guenther et al., 2006), one would hypothesize that talkers with more precise control (i.e., lower variance) would be more sensitive to acoustically perturbed feedback (i.e., larger compensation due to a larger feedback gain parameter). To test this hypothesis, we examined the data from utterances in the meta-analysis that were collected with normal feedback. For each individual, the standard deviation of F1 and F2 was computed from the last 15 utterances of the baseline phase. Initial utterances of the baseline phase were omitted to avoid any potential changes in production as talkers acclimated to talking while wearing headphones. No significant correlation between compensation and standard deviation of normal utterances was found for either F1 or F2 [$r(114) = 0.12$, $p = 0.22$; and $r(114) = -0.08$, $p = 0.41$, respectively].

It is possible that some individuals are unable to independently control F1 and F2. For these individuals, changes in production in F1 and F2 would be linked and thus altering production to compensate in one formant would result in modified production in the other. Thus, these talkers could exhibit a bias in compensating more in one formant than

TABLE I. Average fundamental frequencies, formant frequencies, and durations of seven vowels in an /hVd/ context produced by female talkers with normal feedback. Standard deviation is given in parentheses.

	/i/	/e/	/ɪ/	/ɛ/	/æ/	/ɑ/	/u/
F0 (Hz)	211.3 (12.1)	203.5 (14.1)	205.2 (13.0)	199.8 (14.3)	196.5 (19.0)	196.1 (18.1)	213.4 (12.6)
F1 (Hz)	355.0 (24.8)	462.5 (25.7)	548.4 (26.1)	724.8 (28.9)	903.7 (36.9)	810.1 (29.1)	401.9 (17.4)
F2 (Hz)	2788.0 (54.4)	2620.7 (67.1)	2267.2 (59.6)	2088.5 (45.5)	1836.6 (57.7)	1358.1 (77.7)	1408.6 (84.9)
Duration (ms)	234.9 (27.1)	264.3 (24.5)	177.9 (17.2)	192.6 (20.7)	268.8 (24.3)	290.4 (23.6)	251.3 (23.2)

another. While the results of experiment 1 demonstrated that talkers can modify a single formant, these conclusions were based on results averaged across individuals. Thus, as long as the linkage is not consistent across the population, the results of experiment 1 do not preclude the possibility that for a given individual, the changes in production in F1 and F2 may be linked. To examine this possibility, an analysis of the correlation between F1 and F2 during normal vowel production was conducted. In this analysis, a correlation coefficient was computed for each individual based on 20 utterances collected with normal feedback at the beginning of the experiment. Unlike in experiment 1, the first five utterances with normal feedback were included in the analysis because estimating the steady-state mean value was not the primary focus. Data from the baseline phase for the 48 talkers in experiment 1 was pooled with that of the 116 talkers in the meta-analysis for a total of 164 talkers. Of these talkers, 48 had at least one utterance for which either F1 or F2 could not be estimated so these utterances were omitted in the computation of that talker's correlation coefficient.

A histogram of the correlation coefficient for F1 and F2 across individuals is plotted in Fig. 5. This distribution appears to be normal in shape and has a mean and standard deviation of approximately −0.01 and 0.29, respectively. Here, the correlation coefficient is a measure of the covariance of F1 and F2 between utterances of the same vowel. Thus, a correlation coefficient with a magnitude greater than 0.445 suggests a correlation that is significantly different from zero ($p < 0.05$) and thus exhibits a statistically significant linkage in control of the production of F1 and F2. Conversely, a correlation coefficient of zero (or small magnitude) indicates no linkage in the production of F1 and F2 and suggests each formant is controlled independently. Of the 164 talkers, only 19 (11.6%) exhibited a statistically significant correlation (ten with a positive and nine with a negative correlation). While the results observed here indicate a wide range of individual differences, importantly, the variability observed in this sizeable sample shows no evidence of a relationship between production of F1 and F2 that is consistent across the population. Further, a large majority of the talkers did not exhibit a significant correlation. These
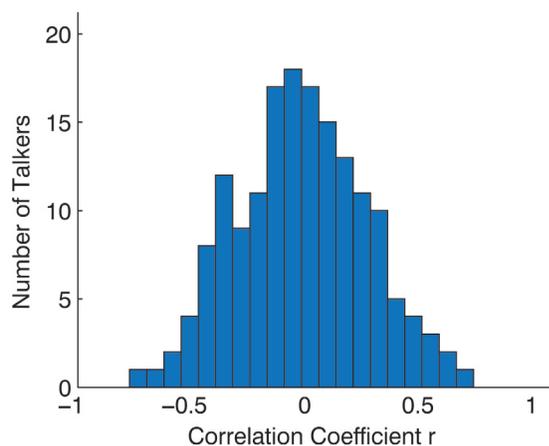

FIG. 5. (Color online) Histogram of individuals' correlation coefficients between F1 and F2 for normal production of /ɛ/ in the word "head."

results are in accordance with a study of talkers imitating steady-state vowel in which some talkers were found to have a significant bias, but the direction of the bias was not consistent across the population (Vallabha and Tuller, 2004).

V. GENERAL DISCUSSION

The results reported in these studies suggest that the system controlling speech production can independently adapt to fine-grained acoustic changes of F1 and F2 but that some modest association in the variance in production of the two formants can be observed. In experiment 1, talkers were able to compensate for local perturbations of a single formant in a manner that did not require a global change to vowel articulation. Thus, the speech motor control system is, in principle, capable of independent control of F1 and F2.

In experiment 2, a large meta-analysis was conducted to examine covariation of the formants. In this analysis, talkers received altered feedback where both F1 and F2 had been perturbed. A moderate correlation was found between individual compensations in F1 and F2. This suggests that the speech motor control system processes the auditory feedback error of F1 and F2 in a common manner at some level. However, in the large meta-analysis of the natural variability in production of normal utterances, most of the population exhibited no significant correlation between the natural variation in production of F1 and F2. No significant correlations were found between individuals' compensations and vowel space differences. Similarly, no significant correlations were found between individuals' compensations and variability in normal production.

The results from experiment 1 address the issue of specificity of control. This issue has a long history in phonetics and phonology (Keating, 1988) and as well in the general study of motor control (Krakauer et al., 2006). The concept of underspecification in phonetics and phonology suggests that some features may not be fully determined in the representation or implemented in the motor realization. This would predict that some aspects of speech motor control would be more tightly constrained while other parts of the vocal tract would exhibit considerable variability. A similar idea has been discussed for articulation of vowels by Perkell and Nelson (1985). They showed that the shape and orientation of xy variance ellipses of tongue tissue points differed for different vowels and that the variance was less at points of maximum constriction in the direction perpendicular to the tongue surface. Shiller et al. (2002) have reported directional differences in kinematic variability of jaw position that were directly related to the biomechanical stiffness of the jaw. Since jaw stiffness can be voluntarily controlled (Shiller et al., 2005), this suggests that speech variability may be regulated in some dimensions more than others. This is consistent with the idea of "minimum intervention" in which only variability that is relevant to the task is tightly controlled (Tresch and Jarc, 2009).

Other studies suggest a somewhat different picture. Nasir and Ostry (2006) reported that small destabilizing force perturbations that do not have any acoustic consequences are compensated for. It appears, thus, that talkers control

MacDonald et al.: Independence of formant control

all aspects of the oral motor system to maintain stability of articulation. In sensorimotor learning contexts, the conditions that support generalization of learning provide insight into the specificity of the sensorimotor representation. Tremblay et al. (2008) trained subjects to adapt to jaw movements in a novel force field. Transfer of learning was not observed to other phonetic contexts even though there was considerable overlap in kinematics. Learning limb movements in a novel force field or in visuomotor transformations similarly show only limited generalization (e.g., Nozaki et al., 2006). These findings support a view in which detailed specific plans are involved in motor control rather than generalized dynamic representations.

Our evidence is consistent with this idea of a detailed representation of articulatory to acoustic mapping by the control system. Perturbations of F2 and, to a certain extent, F1 that raise or lower the formant frequency are strongly compensated for with compensations that are specific to that formant. Studies of force-field perturbations during reaching movements show the same kind of selective compensation (e.g., Shadmehr and Mussa-Ivaldi, 1994). Lateral force-field perturbations during the movement are compensated for without modifying the movements toward the target. Movements thus seem to be decomposable in terms of feedback compensation and in motor learning as well (Flanagan et al., 1999).

While the specificity of control is supported, nothing in our data suggests a similar level of precision of control and this presents a somewhat contradictory picture. As always in speech, the present data show significant variability even though the speech motor system is clearly monitoring the discrepancies between feedback and the intended movement. In experiment 2, we examined the structure of the variability to address the question of independence of control variables in a different manner. Our aim was to test whether the control of the two formants was linked even if the formants could potentially be decoupled in corrective adjustments. A strong correlation between the magnitudes of compensations in F1 and F2 when both are perturbed would suggest a common gain factor for the feedback control system but the modest correlation that was observed indicates that this isn't the case. The origin of this correlation isn't known at this point. Trial-to-trial variability in utterances shows no linkage between the first two formants and the average within-subject correlation is zero.

One of the rationales for synergies in motor control is that a common coordination scheme could be used in various contexts such as different rates, speaking volumes, or linguistic stress levels and a scaling parameter could be used to modify the movement as a whole to fit the context. In limb movements, reaches in the same direction can be learned across different movement extents indicating that gain is separate from the coordination required to plan a movement in the various spatial coordinate frames necessary to reach to a location in extrinsic space (Krakauer et al., 2000). In speech production, there is also evidence for such scalings. Gay's (1974) study of vowel production showed F1/F2 correlations with rate (cf. van Son and Pols, 1992). Scalar changes of the underlying motor organization have been proposed to account for the systematic kinematic changes in

velocity that accompany such rate changes in speech (Ostry and Munhall, 1985). When people speak more clearly, a similar scaling phenomenon can be observed that imparts F1/F2 correlations (Tasko and Greilick, 2010).

While these observed patterns may be attributable to an underlying scaling of a base coordination, they may also be attributable to a common variance source in motor organization. This could include a common noise source and a number of these possible sources of variance in movement have been suggested. They range from motor execution noise (van Beers et al., 2004) to perceptual transduction noise (van Beers et al., 1998) to central planning noise (Churchland et al., 2006). The final motor output variance will be the sum of all of these independent sources and any large single source can have a big impact on movement accuracy. One of the remarkable aspects of the behavior of the motor control system is that it must be able to partial out these sources of variance in order to tell what part is due to random motor noise and what part should be fixed because it is due to programming noise or some system problem like a feedback discrepancy (van Beers, 2009). The tendency of a variable speech motor system to produce corrective movements when feedback is manipulated is evidence for this.

The data provide two negative findings about possible variance relationships associated with compensation magnitude. These negative effects are important, in part, because of the large number of subjects that are used to test the questions. Talkers who have less variable normal productions might be predicted to have larger compensations when perturbed (Perkell et al., 2004; Villacorta et al., 2007). The rationale being that the precision of the articulations by these individuals is due to smaller goal regions for vowels. We found no evidence for this idea when a sample of more than 100 subjects was examined. As noted above, the variance in trial-to-trial speech production is the sum of many independent processes. For the relationship between smaller variability in normal production and larger compensation magnitudes to have held, the compensatory system would have had to be actively predicting and reducing motor error on many levels or the largest component of the variance would have had to have been variation due to the compensatory mechanism. This is clearly not the case.

In addition, the meta-analysis showed that the nearness of adjacent vowels is not related to the compensation magnitude. If the compensatory mechanism was influenced by vowel category boundaries, vowels spaces with more tightly packed vowels might be expected to yield larger compensations to reduce vowel category overlap. This was not the case as there was no correlation between the distance between adjacent vowel formant frequencies and the magnitude of compensation.

In summary, the articulation of vowels shows evidence of a controller that is sensitive to detailed sensory feedback but is also influenced by other sources of variability. The motor control system compensates for feedback perturbations in a specific manner, suggesting a degree of decomposability of the motor plan. The examination of individual differences conducted in the meta-analysis supports this view of the specificity of control. The approach used here of

examining perturbations and also looking at patterns of variation is productive. An optimal control system must be able to accommodate large errors in production and also maintain stability in the face of small errors in natural steady-state conditions (van Beers, 2009). The motor system's solution may be a compromise between quite different constraints. Understanding individual and population patterns will thus require further studies of this kind.

ACKNOWLEDGMENTS

Abbs, J. H., and Gracco, V. L. (**1984**). "Control of complex motor gestures: Orofacial muscle responses to load perturbations of lip during speech," J. Neurophysiol. **51**, 705 723.

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.

Beautemps, D., Badin, P., and Bailly, G. (**2001**). "Linear degrees of freedom in speech production: Analysis of cineradio and labio film data and articulatory acoustic modeling," J. Acoust. Soc. Am. **109**(5), 2165 2180.

Bernstein, N. A. (**1967**). *The Co ordination and Regulation of Movements* (Pergamon Press, Oxford), p. 196.

Browman, C. P., and Goldstein, L. M. (**1986**). "Towards an articulatory phonology," Phonol. Yearbook **3**, 219 252.

Buchaillard, S., Perrier, P., and Payan, Y. (**2009**). "A biomechanical model of cardinal vowel production: Muscle activations and the impact of gravity on tongue positioning," J. Acoust. Soc. Am. **126**(4), 2033 2051.

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**(6), 3153 3161.

Churchland, M. M., Afshar, A., and Shenoy, K. V. (**2006**). "A central source of movement variability," Neuron **52**(6), 1085 1096.

Cowie, R., and Douglas Cowie, E. (**1992**). *Postlingually Acquired Deafness: Speech Deterioration and the Wider Consequences* (Mouton de Gruyter, New York), p. 304.

Diehl, R. L., Lotto, A. J., and Holt, L. L. (**2004**). "Speech perception," Annu. Rev. Psychol. **55**, 149 179.

Flanagan, J. R., Nakano, E., Imamizu, H., Osu, R., Yoshioka, T., and Kawato, M. (**1999**). "Composition and decomposition of internal models in motor learning under altered kinematic and dynamic environments," J. Neurosci. **19**(20), RC34.

Fowler, C. A., and Saltzman, E. (**1993**). "Coordination and coarticulation in speech production," Speech Commun. **36**(2,3), 171 195.

Gay, T. (**1974**). "A cinefluorographic study of vowel production," J. Phonetics **2**, 255 266.

Ghez, C., Favilla, M., Ghilardi, M. F., Gordon, J., Bermejo, R., and Pullman, S. (**1997**). "Discrete and continuous planning of hand movements and isometric force trajectories," Exp. Brain Res. **115**(2), 217 233.

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.

Guenther, F. H., Ghosh, S. S., and Tourville, J. A. (**2006**). "Neural modeling and imaging of the cortical interactions underlying syllable production," Brain Lang. **96**(3), 280 301.

Harshman, R. A., Ladefoged, P., and Goldstein, L. (**1977**). "Factor analysis of tongue shapes," J. Acoust. Soc. Am. **62**, 693 707.

Hoole, P. (**1998**). "Modelling tongue configuration in German vowel production," Proc. 5th Int. Conf. Spoken Lang. Processing **5**, 1863 1866.

Houde, J. F., and Jordan, M. I. (**2002**). "Sensorimotor adaptation of speech I: Compensation and adaptation," J. Speech Lang. Hear. Res. **45**, 295 310.

Howell, P., and Archer, A. (**1984**). "Susceptibility to the effects of delayed auditory feedback," Percept. Psychophys. **36**(3), 296 302.

Jackson, M. T. T. (**1988**). "Analysis of tongue positions: Language specific and cross linguistic models," J. Acoust. Soc. Am. **84**(1), 124 143.

Jones, J. A., and Munhall, K. G. (**2000**). "Perceptual calibration of F0 production: Evidence from feedback perturbation," J. Acoust. Soc. Am. **108**, 1246 1251.

Joos, M. (**1948**). "Acoustic phonetics," Language **24**(2), 5 136.

Kawahara, H. (**1995**). "Hearing voice: Transformed auditory feedback effects on voice pitch control," in *Proceedings of the International Joint Conference on Artificial Intelligence: Workshop on Computational Auditory Scene Analysis*, Montreal, Canada, pp. 143 148.

Kawato, M. (**1989**). "Motor theory of speech perception revisited from the minimum torque change neural network model," in *Eighth Symposium on Future Electron Devices*, Tokyo, Japan, pp. 141 150.

Keating, P. A. (**1988**). "Underspecification in phonetics," Phonology **5**, 275 292.

Kelso, J. A. S., Tuller, B., Vatikiotis Bateson, E., and Fowler, C. A. (**1984**). "Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinate structures," J. Exp. Psychol. Hum. Percept. Perform. **10**, 812 832.

Krakauer, J. W., Mazzoni, P., Ghazizadeh, A., Ravindran, R., and Shadmehr, R. (**2006**). "Generalization of motor learning depends on the history of prior action," PLOS Biol. **4**(10), e316.

Krakauer, J. W., Pine, Z. M., Ghilardi, M. F., and Ghez, C. (**2000**). "Learning of visuomotor transformations for vectorial planning of reaching trajectories," J. Neurosci. **20**(23), 8916 8924.

Ladefoged, P. (**1982**). *A Course in Phonetics*, 2nd ed. (Harcourt Brace, Orlando), p. 300.

Ladefoged, P., and Maddieson, I. (**1996**). *The Sounds of the World's Languages* (Blackwell, Oxford), p. 425.

Lane, H., and Tranel, B. (**1971**). "Lombard sign and role of hearing in speech," J. Speech Hear. Res. **14**, 677 709.

Lee, B. S. (**1950**). "Effects of delayed speech feedback," J. Acoust. Soc. Am. **22**, 824 826.

Lindblom, B. E. F., and Sundberg, E. F. (**1971**). "Acoustical consequences of lip, tongue, jaw, and larynx movement," J. Acoust. Soc. Am. **50**(4), 1166 1179.

Lombard, E. (**1911**). "Le signe de le elevation de la voix" ("The sign of the rise in the voice"), Ann. Malad. l'Orielle Larynx Nez Pharynx (Ann. Dis. Ear, Larynx, Nose Pharynx) **37**, 101 119.

MacDonald, E. N., Goldberg, R., and Munhall, K. G. (**2010**). "Compensations in response to real time formant perturbations of different magnitudes," J. Acoust. Soc. Am. **127**(2), 1059 1068.

Maeda, S. (**1990**). "Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model," in *Speech Production and Speech Modelling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic, Dordrecht), pp. 131 149.

Mooshammer, C., Perrier, P., Fuchs, S., Geng, C., and Payan, Y. (**2001**). "The control of token to token variability: An experimental and modelling study," in *Proceedings of the 4th International Speech Motor Conference*, Nijmegen, pp. 78 81.

Munhall, K. G., Löfqvist, A., and Kelso, J. A. (**1994**). "Lip larynx coordination in speech: Effects of mechanical perturbations to the lower lip," J. Acoust. Soc. Am. **95**(6), 3605 3616.

Munhall, K. G., MacDonald, E. N., Byrne, S. K., and Johnsrude, I. (**2009**). "Talkers alter vowel production in response to real time formant perturbation even when instructed to resist compensation," J. Acoust. Soc. Am. **125**(1), 384 390.

Nasir, S. M., and Ostry, D. J. (**2006**). "Somatosensory precision in speech production," Curr. Biol. **16**(19), 1918 1923.

Nieto Castanon, A., Guenther, F. H., Perkell, J. S., and Curtin, H. D. (**2005**). "A modeling investigation of articulatory variability and acoustic stability during American English /r/ production," J. Acoust. Soc. Am. **117**(5), 3196 3212.

Nix, D. A., Papcun, G., Hogden, J., and Zlokarnik, I. (**1996**). "Two cross linguistic factors underlying tongue shapes for vowels," J. Acoust. Soc. Am. **99**, 3707 3718.

Nozaki, D., Kurtzer, I., and Scott, S. H. (**2006**). "Limited transfer of learning between unimanual and bimanual skills within the same limb," Nat. Neurosci. **9**(11), 1364 1366.

Oller, D. K., and Eilers, R. E. (**1988**). "The role of audition in infant babbling," Child Dev. **59**(2), 441 449.

Orfanidis, S. J. (**1988**). *Optimum Signal Processing, An Introduction* (MacMillan, New York), p. 590.

Ostry, D. J., and Munhall, K. G. (**1985**). "Control of rate and duration of speech movements," J. Acoust. Soc. Am. **77**(2), 640 648.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J., Wilhelms Tricarico, R., and Zandipour, M. (**2000**). "A theory of speech

motor control and supporting data from speakers with normal hearing and with profound hearing loss," J. Phonetics **28**, 233 372.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E. S., and Tiede, M. (**2004**). "The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts," J. Acoust. Soc. Am. **116**, 2338 2344.

Perkell, J. S., and Nelson, W. L. (**1985**). "Variability in production of the vowels /i/ and /a/," J. Acoust. Soc. Am. **77**(5), 1889 1895.

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.

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**(2), 966 977.

Repp, B. H., and Williams, D. R. (**1987**). "Categorical tendencies in imitat ing self produced isolated vowels," Speech Commun. **6**, 1 14.

Saltzman, E. L., and Munhall, K. G. (**1989**). "A dynamical approach to ges tural patterning in speech production," Ecological Psychol. **1**(4), 333 382.

Shadmehr, R., and Mussa Ivaldi, F. A. (**1994**). "Adaptive representation of dy namics during learning of a motor task," J. Neurosci. **14**(5 Pt 2), 3208 3224.

Shaiman, S. (**1989**). "Kinematic and electromyographic responses to pertur bation of the jaw," J. Acoust. Soc. Am. **86**(1), 78 88.

Shiller, D. M., Houle, G., and Ostry, D. J. (**2005**). "Voluntary control of human jaw stiffness," J. Neurophysiol. **94**(3), 2207 2217.

Shiller, D. M., Laboissière, R., and Ostry, D. J. (**2002**). "Relationship between jaw stiffness and kinematic variability in speech," J. Neurophy siol. **88**(5), 2329 2340.

Tasko, S. M., and Greilick, K. (**2010**). "Acoustic and articulatory features of diphthong production: A speech clarity study," J. Speech Lang. Hear. Res. **53**(1), 84 99.

Ting, L. H. (**2007**). "Dimensional reduction in sensorimotor systems: A framework for understanding muscle coordination of posture," in *Compu tational Neuroscience: Theoretical Insights into Brain Function, Progress in Brain Research*, Vol. 165, edited by P. Cisek, T. Drew, and J. F. Kalaska, (Elsevier, Amsterdam), pp. 301 325.

Ting, L. H., and McKay, J. L. (**2007**). "Neuromechanics of muscle synergies for posture and movement," Curr. Opin. Neurobiol. **17**(6), 622 628.

Tremblay, S., Houle, G., and Ostry, D. J. (**2008**). "Specificity of speech motor learning," J. Neurosci. **28**(10), 2426 2434.

Tresch, M. C., and Jarc, A. (**2009**). "The case for and against muscle syn ergies," Curr. Opin. Neurobiol. **19**(6), 601 607.

Vallabha, G. K., and Tuller, B. (**2004**). "Perceptuomotor bias in the imita tion of steady state vowels," J. Acoust. Soc. Am. **116**(2), 1184 1197.

van Beers, R. J. (**2009**). "Motor learning is optimally tuned to the properties of motor noise," Neuron **63**, 406 417.

van Beers, R. J., Haggard, P., and Wolpert, D. M. (**2004**). "The role of execution noise in movement variability," J. Neurophysiol. **91**(2), 1050 1063.

van Beers, R. J., Sittig, A. C., and Denier van der Gon, J. J. (**1998**). "The precision of proprioceptive position sense," Exp. Brain Res. **122**(4), 367 377.

van Son, R. J. J. H., and Pols, L. C. W. (**1992**). "Formant movements of Dutch vowels in a text, read at normal and fast rate," J. Acoust. Soc. Am. **92**(1), 121 127.

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**(4), 2306 2319.

Wolpert, D. M., Ghahramani, Z., and Flanagan, J. R. (**2001**). "Perspectives and problems in motor learning," Trends Cogn. Sci. **5**(11), 487 494.

Wood, S. (**1982**). *X Ray and Model Studies of Vowel Articulation*, Working Paper 23, Department of Linguistics, Lund University, p. 192.

Yates, A. J. (**1963**). "Delayed auditory feedback," Psychol. Bull. **60**(3), 213 232.