A cross-language study of compensation in response to real-time formant perturbation

Takashi Mitsuya[a] and Ewen N. MacDonald[b]
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[c]
Department of Psychology, Queen's University, Humphrey Hall, 62 Arch Street, Kingston, Ontario K7L 3N6, Canada

(Received 1 February 2011; revised 24 August 2011; accepted 5 September 2011)

Past studies have shown that when formants are perturbed in real time, speakers spontaneously compensate for the perturbation by changing their formant frequencies in the opposite direction to the perturbation. Further, the pattern of these results suggests that the processing of auditory feedback error operates at a purely acoustic level. This hypothesis was tested by comparing the response of three language groups to real-time formant perturbations, (1) native English speakers producing an English vowel /ɛ/, (2) native Japanese speakers producing a Japanese vowel (/e̞/), and (3) native Japanese speakers learning English, producing /ɛ/. All three groups showed similar production patterns when F1 was decreased; however, when F1 was increased, the Japanese groups did not compensate as much as the native English speakers. Due to this asymmetry, the hypothesis that the compensatory production for formant perturbation operates at a purely acoustic level was rejected. Rather, some level of phonological processing influences the feedback processing behavior.
 [DOI: 10.1121/1.3643826]

PACS number(s): 43.70.Mn, 43.70.Kv, 43.70.Bk [AL] Pages: 2978–2986

I. INTRODUCTION

Evidence from a number of sources supports the idea that speakers constantly monitor the sounds they produce so that the produced outcome is consistent with what was intended. This self-production–perception relationship has been substantiated by a variety of studies. For example, clinical studies have reported more variable articulation among postlingually deafened and hearing-impaired individuals (Waldstein, 1990; Cowie and Douglas-Cowie, 1992; Schenk *et al*., 2003). Laboratory studies further examined this perception–production linkage by perturbing auditory feedback in real time. In this paradigm, speakers produce a speech segment repeatedly, and some of the acoustic parameters are modified so that the feedback is incongruent with the intended articulation. These studies have shown that speakers spontaneously compensate not only for suprasegmental manipulations, such as loudness (Bauer *et al*., 2006) and pitch (Burnett *et al*., 1998; Jones and Munhall, 2000), but also for segmental manipulations such as vowel formant frequency (Houde and Jordan, 2002; Purcell and Munhall, 2006; Villacorta *et al*., 2007) and fricative acoustics (Shiller *et al*., 2009).

In many of the segmental manipulation studies using this paradigm, the formant structure of a vowel was manipulated in real time while speakers produced a simple word like "head" (Purcell and Munhall, 2006; MacDonald *et al*., 2010; MacDonald *et al*., 2011). With altered formant structure, speakers heard themselves say a slightly different vowel than the one intended. For example, when the first formant (F1) of /ɛ/ for the word head was increased in frequency by 200 Hz, the speakers' feedback sounded more like "had" (/hæd/). In response to this altered feedback, speakers spontaneously lowered the frequency of F1, so that the feedback was more consistent with the intended vowel. A similar behavior was observed when the vowel's F1 was decreased in frequency; subjects produced compensations in the opposite direction in frequency to the perturbation.

The magnitude of compensation to perturbations for these two directions of shift was reported to be similar, although average compensations were smaller than the magnitude of perturbation applied. The relationship between magnitude of perturbation and compensation was closely examined by a recent study (MacDonald *et al*., 2010). In this study, gradual, incremental perturbations were applied to the vowel formant structure, and the compensatory magnitude was reported to be a constant proportion of the applied perturbation (i.e., proportional compensation at ~25%–30%).

[a]Author to whom correspondence should be addressed. Electronic mail: takashi.mitsuya@queensu.ca
[b]Also at: Center for Applied Hearing Research, Department of Electrical Engineering, Technical University of Denmark, Ørsteds Plads, Building 352, DK-2800 Kgs. Lyngby, Denmark.
[c]Also at: Department of Otolaryngology, Queen's University, Humphrey Hall, 62 Arch Street, Kingston, Ontario K7L 3N6, Canada.

The symmetrical response to perturbations that either raise or lower formant frequency and the approximate linearity of the compensatory patterns across a wide range of perturbation magnitudes strongly suggest that articulator trajectory control is regulated without the transformations involved in phonetic perception. If phonemic organization is involved in the processing of error feedback, then compensation should be sensitive to phonemic category boundaries and result in a nonlinear relationship between compensation and perturbation magnitudes.

In principle, one might test the hypothesis that compensatory production operates at a purely acoustic level without phonological mediation by testing different vowels within a language. However, vowels from some regions of the vowel space might have different acoustic–articulatory relations [e.g., quantal effects as proposed by Stevens (1989)]. Somatosensory feedback also plays a role in the control of speech production and this feedback may vary across the vowel space. Vowels like /i/ have lingual contact with the teeth and palate and more open vowels have strong joint and muscle sense information for articulator position. Thus, if differences in compensatory response across vowels were observed, phonological constraints and sensorimotor differences would be confounded. In contrast, cross-language tests of vowels allow measures of vowels in similar regions of the vowel space with significant differences in the native language phonologies.

Languages differ in the number, location, and relative proximity of vowels, and thus each language presents different articulatory challenges. For example, requirements for the precision of articulation may vary between languages with closely vs sparsely distributed vowel inventories (e.g., Manuel, 1990). Further, languages may also vary in syllabic structure, vowel quality, and temporal contrasts. Thus, the requirements for successful control of formant production may vary across language. All of these phonological characteristics, however, should not influence a low-level acoustic-articulator mapping. If feedback processing is influenced by the phonology itself then cross-language differences in the pattern of compensation can be expected.

To our knowledge, no cross-language formant perturbation studies have been reported in the literature. Although some work has been done with native Mandarin speakers (e.g., Cai *et al.*, 2010), the majority of the formant perturbation studies have focused on native English speakers producing an English vowel. Thus, the compensatory production for monophthongs reported in previous studies might have been unique to (1) the participants' producing their native vowel and/or (2) English vowels. Whether or not the same compensatory pattern is observed with non-native vowels or with other language speakers producing their native vowel is unknown. By conducting a cross-language examination, we can confirm or reject the hypothesis that the acoustic feedback error is processed at a purely acoustic level. If different patterns are observed, it would imply that phonological processes are influencing the processing of error signals in the control of production.

Because past studies' findings regarding monophthongs are limited to (1) production of native vowels and (2) the English vowel inventory, these factors were considered in the current study. Specifically, the level of language experience and native vowel space were considered by examining the response to auditory feedback perturbations in native Japanese speakers speaking Japanese and Japanese speakers of English as a second language (ESL).

The sounds of a second language (L2) or an unfamiliar language are not perceived and produced precisely if they are not part of one's native language (L1) sound inventory. Perceptually, at an early stage of learning a new language, the L2 sounds are categorized into speakers' L1 sound categories, but not consistently. For example, Japanese ESL speakers tend to categorize the English vowel /ɪ/ as the Japanese /i/ or /ii/ most of the time, yet they may also categorize the vowel as the Japanese /e̞/, /e̞ e̞/, or /e̞ i/ (Strange *et al.*, 1998; Strange *et al.*, 2001). Similarly, a study by Nishi and Kewley-Port (2007) showed that when asked to identify the English vowel /ɪ/, Japanese ESL speakers reported the English vowel /ɛ/ most of the time.

Production studies also have reported large variability for L2 vowels (Chen *et al.*, 2001; Ng *et al.*, 2008; Wang and van Heuven, 2006). These studies reported that ESL learners' English vowel spaces not only had slightly different acoustic targets but also larger variability in formant production. The larger variability of L2 vowels may be because acoustic and/or articulatory targets with non-native language sounds are not well defined, motor coordination to achieve such targets is more variable, or a combination of the two. Regardless of the source of variability, the production of speech sounds differs in the level of practice and experience in producing the sounds between one's L1 and L2. Due to this inherent difference of productive experience, it is possible that how people compensate for formant-shifted feedback may depend on whether they are producing the sounds of L1 vs L2 during the experiment.

If the level of experience in producing sounds has an effect, ESL speakers of Japanese should show a different compensatory production, depending on whether they are producing their native Japanese or non-native English vowels, whereas the pattern of the two language groups producing their L1 vowel should not differ. If, however, the behavior is solely due to a frequency-based articulatory mapping adjustment and no linguistic processes are involved, then there should not be a difference across any of the conditions.

If there is no difference in compensatory production between native English speakers and Japanese ESL speakers, this would suggest that there is no level of experience effect. However, there still might be a difference between Japanese L1 compensatory production and that of English speakers' L1 due to language specific vowel phonology. Japanese and English differ on such factors as differences in acoustic target, vowel density, durational cues, syllabic structure, suprasegmental cues and so on. These intrinsic differences between Japanese and English may impose different requirements on the use of feedback in speech production control. Given that formant perturbation manipulates the quality of the vowel being produced, feedback processing may be specific to the particular vowel phonology.

In the current study we compared compensatory production of three groups: (1) native English speakers producing English /ɛ/ (ENG L1), (2) native Japanese speakers producing Japanese /e̞/ (JPN L1), and (3) native Japanese speakers producing English /ɛ/ (JPN L2). Briefly, if vowel quality and distribution of vowels affect speech control, then the difference in compensatory production with formant-shifted feedback is expected between ENG L1 and JPN L1 groups. However, if only language experience affects the behavior, we would expect that there might be a difference between ENG L1 and JPN L2, but no difference between ENG L1 and JPN L1 groups. Last, if the compensatory production operates at a purely acoustic level, without any mediation of phonological processing, then these three groups should not show any differences.

II. EXPERIMENT

A. Participants

The participants in this study consisted of 18 native English speakers and 35 native Japanese speakers. All speakers were female. The study was restricted to one gender to reduce formant differences across participants. The English speakers were undergraduate students at Queen's University and their average age was 19.38 (ranging from 18 to 23 years). The Japanese speakers were students studying English as a second language at Queen's School of English. With the exception of two participants, all the Japanese speakers had been in Canada for less than 8 months at the time of the study. The majority of Japanese speakers was enrolled in the intermediate level of the ESL program, and had not been immersed in an English speaking culture prior to coming to Canada. The other two Japanese speakers were undergraduate students at Queen's University. One had finished the highest level of the ESL program at Queen's University and the other had moved to Canada from Japan during her high school years. The mean age of the Japanese group was 21.3 years (ranging from 19 to 32 years). Each participant was tested in a single session. No participants reported speech or language impairments and all had normal audiometric hearing thresholds over a range of 500–4000 Hz.

B. Equipment

Equipment used in this study 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) and wore a headset microphone (Shure WH20) and the microphone signal was amplified (Tucker-Davis Technologies MA3 microphone amplifier), low-pass filtered with a cutoff frequency of 4.5 kHz (Krohn-Hite 3384 filter), digitized at 10 kHz and filtered in real time to produce formant shifts (National Instruments PXI-8106 embedded controller). 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 ~80 and 50 dBA SPL, respectively.

C. 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 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.

D. 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 autoregressive analysis. Prior to data collection, speakers produced six utterances of seven English vowels, /i, ɪ, e, ɛ, æ, ɔ, and u/, in an /hVd/ context ("heed," "hid," "hayed," "head," "had," "hawed," and "who'd"). When a word appeared on a screen in front of them, speakers were instructed to say the prompted word without gliding the tone or pitch.

For the Japanese participants, the five Japanese vowels /a, e̞, i, o, and ɯ/ were also collected in a similar manner after collecting the English vowels. Participants were asked to produce the Japanese vowels in an /hV/ context. The words were shown on the monitor in Hiragana ("は", "へ", "ひ", "ほ", "ふ"). Similar to the English words, these Japanese prompts were lexical items in Japanese.

For both English and Japanese speakers, utterances were analyzed with model orders ranging from 8 to 12. For each talker, 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 (MacDonald *et al.*, 2010).

E. Offline formant analysis

The procedure used for offline formant analysis was the same 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. Although 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 for each individual.

F. Procedure

The participants were split into three experimental groups. The 18 native English speakers were assigned to the English L1 group (ENG L1). The native Japanese speakers were randomly assigned to either the JPN L1 ($n = 17$) or JPN L2 ($n = 18$) group. The speakers in these two groups did not differ in the time they had spent in Canada prior to the study. Other than the utterances produced for the experiment, the procedure for each group was identical. Both the ENG L1 and JPN L2 groups produced utterances of the English word "head" (/hɛd/), whereas the JPN L1 group produced utterances of the Japanese word "へ" (/he̞/).

The experiment consisted of two conditions: F1 Increase and F1 Decrease. In the F1 Increase condition, when feedback was altered, F1 was increased. Similarly, in the F1 Decrease condition, when feedback was altered, F1 was decreased. The order in which speakers performed the F1 Increase and F1 Decrease conditions was counterbalanced and speakers received a short break between conditions.

Over the course of each condition, each speaker produced 140 utterances. Each condition consisted of four phases (Fig. 1). In the first phase, Baseline (utterances 1–20), speakers received normal feedback (i.e., amplified and with noise added but with no shift in formant frequency). In this and subsequent phases, subjects were encouraged to speak at a natural rate and 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. In the second phase, Ramp (utterances 21–70), speakers produced utterances while receiving altered feedback in which F1 was increased (F1 Increase condition) or decreased (F1 Decrease condition) by 4 Hz per utterance. In the third phase, Hold (utterances 71–90), speakers received altered feedback in which F1 was increased (F1 Increase condition) or decreased (F1 Decrease condition) by 200 Hz. In the final phase, Return (utterances 91–140), utterances were produced with normal feedback (i.e., the formant shift was abruptly turned off).

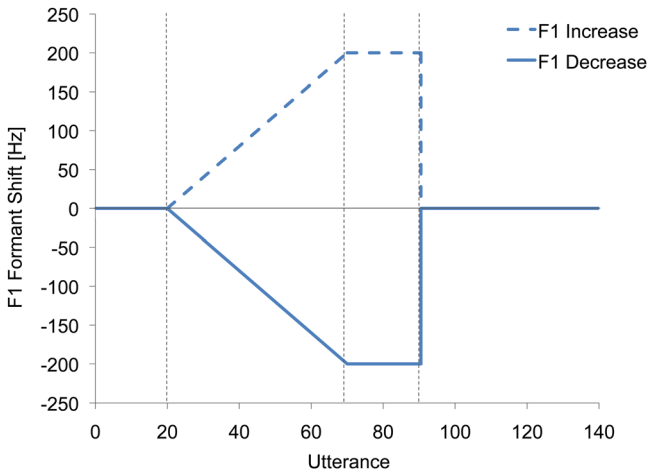

FIG. 1. (Color online) Feedback shift applied to the first formant for the F1 Decrease (solid line) and F1 Increase (dashed line) conditions. The vertical dashed lines denote the boundaries of the four phases: Baseline, Ramp, Hold, and Return.

III. RESULTS

The baseline average of F1 was calculated from the last 15 utterances of the Baseline phase (i.e., utterances 6–20) and the F1 results were then normalized by subtracting the subject's baseline average from each utterance value. The normalized results for each utterance, averaged across speakers, can be seen in Fig. 2. All three groups of speakers compensated for the altered feedback in both shift conditions by changing production of F1 in a direction opposite that of the perturbation. However, differences in the magnitude of the response were observed across groups.

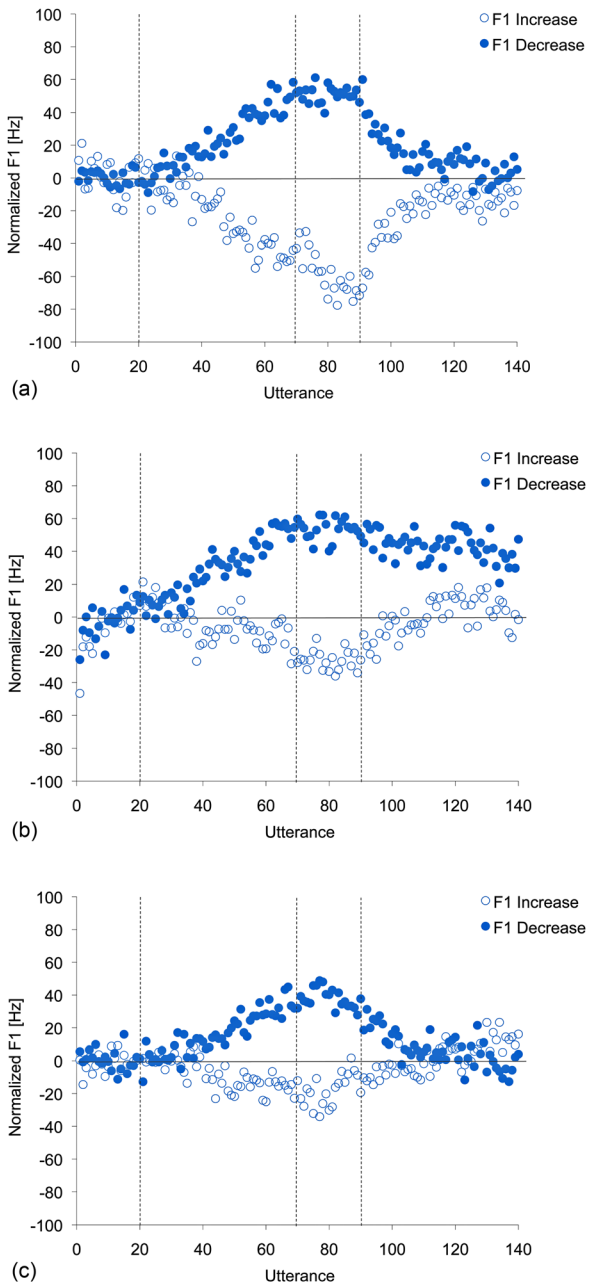

FIG. 2. (Color online) Normalized F1 production averaged across talkers for the F1 Decrease (solid circles) and F1 Increase (open circles) condition for (A) ENG L1, (B) JPN L1, and (C) JPN L2 groups. The vertical dashed lines denote the boundaries of the four phases: Baseline, Ramp, Hold, and Return.

To quantify the change in production, we defined compensation as the magnitude of the change in formant frequency from the baseline average with sign based on whether the change opposed (positive) or followed (negative) that of the perturbation. A measure of average compensation was computed by averaging over the utterances of the Hold phase (i.e., utterances 71–90) for each individual in each shift condition. The average compensation of each group is plotted in Fig. 3. For each of the shift conditions, an analysis of variance (ANOVA) was conducted on the average compensation data with Order of shift condition and Language groups as between-subject factors. For the F1 Decrease condition, none of the effects was significant, indicating that neither order nor language group had any effect on the compensation magnitude. However, for the F1 Increase condition, only a significant main effect of Language group was found [$F(2, 47) = 8.215$, $p < 0.001$]. *Post hoc* analyses with Bonferroni correction showed that the speakers in the ENG L1 group produced significantly larger compensation compared to JPN L1 ($p < 0.05$) and JPN L2 speakers ($p < 0.01$). However, the two Japanese groups did not differ ($p > 0.10$) from each other.

In order to examine the symmetry of the magnitude of compensation, paired sample *t*-tests were performed within each group, using the average magnitude of compensation for the last 20 trials in the Hold phase. Although both ENG L1 and JPN L2 groups showed no significant results, JPN L1 group's difference approached significance [$t(16) = -1.915$, $p = 0.07$], suggesting a slightly larger compensation in the F1 Decrease condition [$X = 53.27$; standard deviation (s.d.) = 53.39] than the F1 Increase condition ($X = 26.76$; s.d. = 26.52).

Although F2 was, in general, more variable than that of F1, overall, our speakers did not change their F2 production during the Hold phase. The change in production of F2 was examined in the same way as F1 production. Using the average F2 during the last 15 utterances of the baseline as a reference, speakers' average change in F2 production during the Hold phase was computed. A mixed design ANOVA with language groups as a between-subjects and direction of F1 shift as a within-subjects factor was conducted. No significant interaction or main effect of either factor was observed ($p > 0.05$).

In the Return phase, de-adaptation in the F1 Decrease condition among the JPN L1 group displayed a different pattern from that of the other groups, showing that the compensatory production seemed to be maintained even after the perturbation was removed. A one-way ANOVA with Language groups as a between subject variable was performed on the average magnitude of compensation for the last 20 trials in the Return phase of the F1 Decrease condition, and it revealed a significant group difference [$F(1,2) = 4.97$, $p < 0.05$], and the subsequent *post hoc* analysis with Bonferroni correction (α set at 0.016) confirmed that the JPN L1 group in the F1 Decrease condition was significantly larger than that of ENG L1 and JPN L2 groups ($p < 0.016$), whereas these two groups did not differ significantly ($p > 0.016$). In the F1 Increase condition, on the other hand, there was no group difference ($p > 0.05$).

The quality of the vowels produced during the Baseline phase was examined to verify that the vowels produced by each group were spectrally similar. For each individual, the average F1 and F2 were calculated from the last 15 utterances of the Baseline phase (i.e., utterances 6–20). The average baseline F1 for each group can be seen in Fig. 4. A repeated measures ANOVA was conducted for the average baselines of F1 and F2, with Shift Condition as a within and with Language Group and Order of shift direction as between-subject factors. For F1, although the average F1 of the utterances produced by the native Japanese speakers was slightly lower than those of the native English speakers, no significant interactions or main effects were found (all with $p > 0.05$). However, for F2 (see Fig. 5), a significant main effect of group was found [$F(1,2) = 34.584$, $p < 0.001$]. The *post hoc* analysis with Bonferroni adjustment showed that the two Japanese groups showed significantly higher value of F2, compared to the ENG L1 group, indicating that the vowel

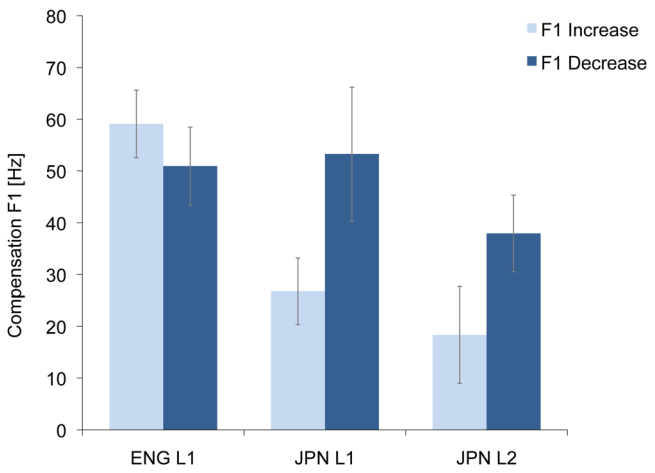

FIG. 3. (Color online) Average compensation in F1 over the Hold phase (i.e., utterances 71–90). Compensation is defined as the magnitude of the change in formant frequency from the baseline average with sign based on whether the change opposes (positive) or follows (negative) that of the perturbation. Error bars indicate one standard error.

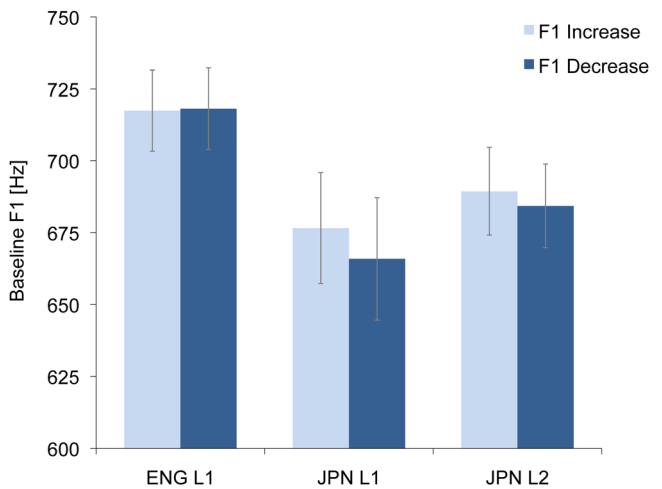

FIG. 4. (Color online) Average F1 computed from the last 15 utterances of the Baseline phase (i.e., utterances 6–20). Error bars indicate 1 standard error.

that our Japanese speakers were producing was more fronted than the one that the ENG L1 speakers produced.

The difference in F2 produced across groups led us to speculate that the cross-language differences in compensation observed in this study may be the result of inherent differences in producing vowels with slightly different positions in the vowel space. The Japanese groups produced vowels with a higher F2 value, and in this particular location of the vowel space, it may be more difficult to alter production in a manner that compensates for perturbations that increase F1. If so, native English speakers who have a relatively higher baseline F2 frequency for /ɛ/ should also compensate in a similar manner to Japanese speakers. To examine this, the ENG L1 group was split into two subgroups using a median split based on the average F2 value of /ɛ/ collected during the prescreening procedure. A mixed design ANOVA was performed with the F2 grouping as a between-subject and Shift directions as a within-subject factor. None of the effects were significant ($p > 0.05$). Moreover, we also examined the correlation between the magnitude of compensation and speakers' average baseline F2 value. For both shift directions, no significant correlation was found among all of the groups (all $p > 0.05$). The observation of no relationship between the F2 value and magnitude of compensation within each group suggests that the difference in compensatory pattern between ENG L1 vs Japanese groups is not likely a result of modest cross-group differences in formant acoustics.

To examine the precision of control of vowel production, the standard deviation of F1 was calculated for each speaker based on the last 15 utterances of the Baseline phase. A repeated measures ANOVA, with the Order of shift direction as a within-subject and Language group as a between-subject factor was carried out. None of the effects were significant ($p > 0.05$), indicating that the production of the vowels, regardless of the speakers' native languages and L1/L2 status, was relatively stable across the groups and the shift conditions. Thus, no statistically significant differences in the quality or stability of vowel production were observed across groups. For F2, an ANOVA revealed that JPN L2 group had significantly larger variability during the baseline phase, compared to the ENG L1 group [JPN L2: mean = 58.0, standard error (s.e.) = 7.6; ENG L1: mean = 40.4, s.e. = 3.3]. However, the variability of the JPN L1 (mean = 48.9,

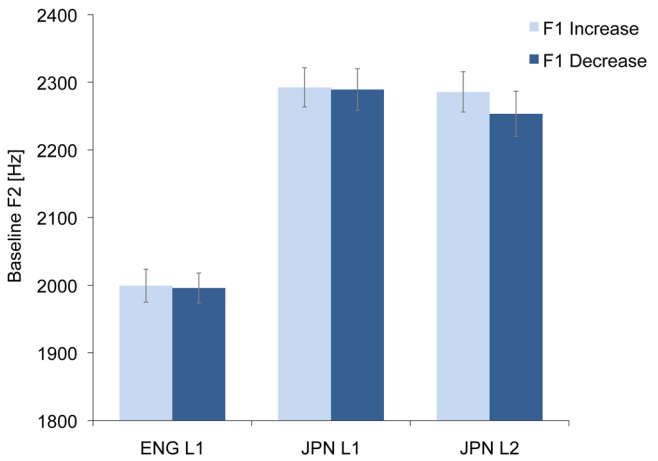

FIG. 5. (Color online) Average F2 computed from the last 15 utterances of the Baseline phase (i.e., utterances 6–20). Error bars indicate one standard error.

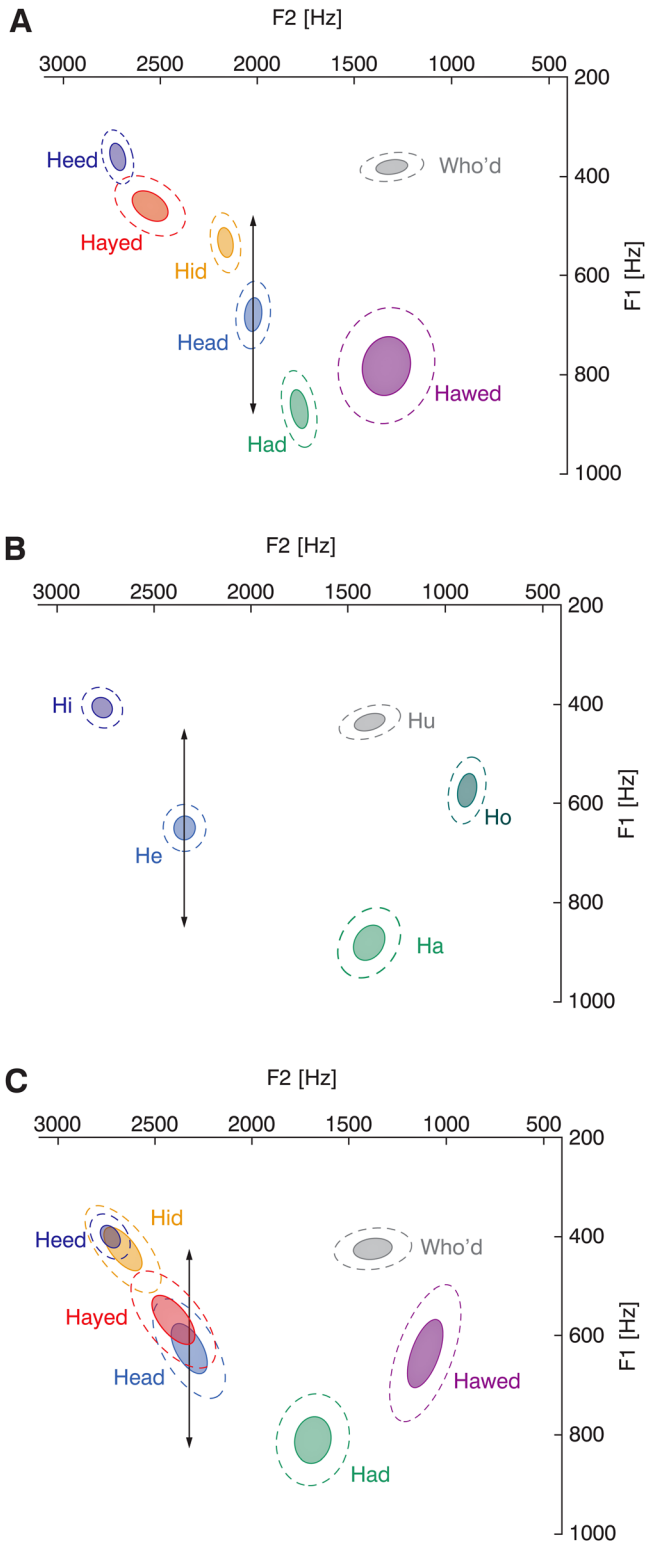

FIG. 6. (Color online) Vowel space of an average individual speaker from (A) ENG L1, (B) JPN L1, and (C) JPN L2 groups. The centroid of each ellipse represents the average F1/F2 value for that vowel. The solid and dashed ellipses represent one and two standard deviations, respectively. The arrows indicate the 200 Hz perturbation and the arrow tips indicate what the acoustic feedback speakers would have heard if they did not compensate.

s.e. = 4.2) group did not differ significantly from that of either JPN L2 or ENG L1 groups.

The native language and second language vowel spaces may vary in the distribution of vowels and this might influence compensation. Vowel spaces were estimated based on the words collected for estimating model order. The seven English vowels in an /hVd/ context (heed, hid, hayed, head, had, hawed, and who'd) were used to estimate the English vowel space of our ENG L1 and JPN L2 groups. Similarly, for Japanese speakers, estimates of the Japanese vowel space were calculated from the five Japanese vowels produced in an /hV/ context. The results for an average individual speaker are plotted in Fig. 6. The results for the English and Japanese vowels are also found in Tables I and II, respectively. In comparing the vowel spaces, we see that the distribution of front English vowels produced by JPN L2 is strongly assimilated to their native Japanese vowels, in such a way that /ɪ/ is produced almost identical to /i/ and /ɛ/ is similar to /e̞/.

IV. DISCUSSION

The current study examined cross language differences in the compensatory production for real-time formant perturbations. The results for our native English speakers replicated previous findings that when F1 of the vowel /ɛ/ was increased or decreased in frequency, the speakers compensated for the change by altering the frequency of F1 in the direction opposite the shift (Purcell and Munhall, 2006; Villacorta *et al.*, 2007; MacDonald *et al.*, 2011). Moreover, the magnitude of compensation for both shift directions was very similar. For the F1 Decrease condition, our Japanese speakers also compensated for the formant perturbation in a similar manner to English speakers regardless of whether they were producing an English /ɛ/ or Japanese /e̞/. However, in the F1 Increase condition, the Japanese speakers exhibited significantly less compensation compared to the native English speakers in the same condition.

This asymmetry in the magnitude of compensation by Japanese speakers might have a number of different explanations. The origin of the difference may have something to do with speech motor coordination. Speakers of different languages may have varying levels of experience in producing the gestures for vowels in a given location in the vowel space. Thus, when presented with formant perturbations, speakers of one language may be less familiar with producing the required compensatory gestures, regardless of the manner in which the acoustic feedback error is processed. For example, native Japanese speakers might have less experience in producing compensatory gestures close to /æ/ than speakers of English since this vowel requires movements outside of the normal Japanese vowel space. However, it was not gestures in this region that produced the observed differences between Japanese and English. The cross-language difference in the compensatory response that we observed was for gestures that were comfortably within (e.g., toward the center of) the vowel space rather than out of the vowel space. Thus, it is unlikely that the cross-language pattern observed here is a result of differing abilities to produce vocal tract movements.

The observed asymmetry of compensatory production thus appears to be due to cross-language differences in how auditory feedback is perceptually processed and applied to formant control. When the acoustic feedback is perturbed outside of the acoustic Japanese vowel space (Vance, 1987), less compensation was observed than that produced by English speakers. When the formant feedback is perturbed into the acoustic Japanese vowel space, compensation by Japanese speakers was similar to that of English speakers.

The cross-language differences in compensation are consistent with the perception of auditory feedback error being influenced by the distribution of vowels in the vowel space (see Fig. 6). For native English speakers producing the vowel /ɛ/, the acoustic consequence of both the F1 Increase and F1 Decrease shifts would be feedback that was close to the acoustic region of an adjacent vowel (/æ/ and /ɪ/ respectively).

Similarly, for Japanese speakers producing /e̞/, the acoustic consequence of the F1 Decrease shift would be feedback that was close to the Japanese /i/. Thus, speakers may have compensated in order to maintain the perceptual distinctiveness of the vowel produced (English /ɛ/ or Japanese /e̞/). For Japanese speakers producing /e/, the acoustic consequence of the F1 Increase shift would be feedback that was similar in F1 to the Japanese /a/. However, the Japanese vowel /a/ has a much lower F2. Although the native Japanese speakers may not have perceived the perturbed feedback as a good token of /e/, they may have perceived it as being an acceptable token. Thus, there would be less need to compensate to maintain perceptual distinctiveness. Taken together, we can reject the hypothesis that compensatory production is purely a frequency-based error reduction process. Instead, we hypothesize that the nature of acoustic feedback is phonologically mediated. This

TABLE I. Acoustic parameters of English vowels. An average individual's mean F1 and F2 (with standard deviation in parentheses) of the seven English vowels produced in an /hVd/ context for each group.

	/i/		/ɪ/		/e/		/ɛ/		/æ/		/ɔ/		/u/	
	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
ENG	365.09	2694.97	529.90	2146.87	465.69	2516.92	685.50	2009.86	877.49	1769.82	784.26	1345.34	384.69	1293.44
L1	(20.1)	(40.6)	(21.6)	(34.8)	(25.0)	(78.6)	(29.1)	(39.3)	(33.7)	(45.7)	(38.1)	(72.8)	(14.8)	(80.8)
JPN	394.25	2749.47	416.27	2684.46	547.08	2489.22	638.07	2319.80	778.73	1792.11	611.76	1074.98	421.97	1473.73
L2	(18.2)	(52.4)	(29.6)	(64.5)	(31.8)	(81.3)	(38.7)	(62.1)	(46.0)	(76.3)	(44.9)	(86.9)	(21.3)	(89.8)

TABLE II. Acoustic parameters of Japanese vowels produced by each group of Japanese speakers. An average individual's mean F1 and F2 (with standard deviation in parentheses) of the five Japanese vowels produced in an /hV/ context for each group.

	/i/		/e̞/		/a/		/o/		/ɯ/	
	F1	*F2*	*F1*	*F2*	*F1*	*F2*	*F1*	*F2*	*F1*	*F2*
JPN L1	411.38	2734.08	668.23	2334.95	895.23	1419.57	593.11	905.57	435.25	1363.05
	(17.6)	(41.5)	(21.8)	(45.4)	(29.0)	(65.5)	(34.2)	(52.5)	(13.7)	(74.5)
JPN L2	399.96	2765.79	614.55	2374.12	840.95	1369.97	533.35	869.27	437.47	1506.88
	(22.7)	(54.0)	(29.3)	(58.3)	(39.2)	(74.9)	(27.1)	(36.9)	(20.2)	(87.7)

interpretation is consistent with the findings reported by Manuel (1990) for extent of coarticulation as a function of vowel space density.

The nature of the phonological mediation is at present not clear. Previous studies (e.g., Purcell and Munhall, 2006; MacDonald *et al.*, 2010) have shown that speakers initiate compensation for relatively small perturbations (<100 Hz for F1) and make approximately linear compensations to feedback changes that clearly cross a category boundary without producing any inflection in the response. A number of aspects of vowel representation may be playing a role in these results including category prototypes, dimensions of category goodness, perceptual ambiguities relative to adjacent vowels and various nonlinearities of the vowel perceptual space (e.g., Kuhl, 1991).

In order to further examine the relationship between phonological category and compensatory behavior, it will be important to examine speakers' vowel perception as well as their articulatory behavior. The vowel spaces obtained from the current study are only measures of production. Although they may be related, the ellipses plotted in Fig. 6 represent variability of production, and do not necessarily indicate speakers' perceptual categorical boundaries or goodness ratings.

Another possible explanation to account for the current data is the Natural Referent Vowel (NRV) perspective. Polka and Bohn (2003) found that discrimination between two vowels is better when the presentation of a more central vowel is preceded by a peripheral referent vowel. This asymmetry in perceptual saliency is indeed comparable to our data pattern observed. In the case of our English speakers, in both shift directions, more central vowel /ɛ/ was shifted toward the functionally more peripheral referent vowels /ɪ/ and /æ/, whereas with our Japanese speakers, in the F1 Increase condition, the perturbation shift was far from the reference vowel /a/. Further investigation is needed to examine this account.

Unfortunately, the current data do not allow us to draw a strong conclusion on the effect of language experience (i.e., native vs non-native language status). Although a difference was observed between the English and JPN L2 group, no difference was observed between the two Japanese groups. It is possible that Japanese speakers in the L2 group produced a substituted L1 vowel. Indeed, the similarity in the baseline vowel production both in terms of formant structure and variability between the JPN L1 and L2 groups supports this possibility. However, it is also possible that the /ɛ/ produced by the JPN L2 group was strongly assimilated toward Japanese /e̞/. During the prescreening procedure, our Japanese speakers did differentiate the production of F2 for the English /ɛ/ and the Japanese /e̞/. Thus, there is evidence that at some level, these vowels are distinct in their repertoire. In this phase of the experiment, multiple vowels were produced, which may have resulted in our Japanese speakers making a conscious effort to maintain vowel contrasts. During the experimental trials in which a single vowel was produced repetitively, a less familiar L2 vowel might have been produced in a manner that is more similar to a familiar native vowel. If our JPN L2 speakers were trying to produce the English /ɛ/ (with articulation that was very similar to the Japanese /e̞/) then our results would suggest that there was a language experience effect. However, if JPN L2 speakers were simply substituting the Japanese /e̞/ when producing the English /ɛ/, then the level of language experience was not properly tested.

In addition to a difference in compensation magnitude for the F1 Increase condition, some other differences in the time-course of compensation were observed across the three groups. For the main analysis, we defined the magnitude of compensation by averaging change in F1 production over the Hold phase. However, the stability of compensatory production during this phase appeared to be different across the language groups as well as the shift directions and this difference influenced the magnitude measures. Although native English speakers exhibited a strong and relatively stable compensation in both shifting directions, the Japanese groups showed different patterns. In the F1 Increase condition, the JPN L1 speakers, and in both shifting conditions for the JPN L2 speakers, it seems that the speakers stopped compensating in the middle of the Hold phase even though the maximum perturbation was still being applied. Their F1 production drifted back somewhat to the normal baseline. During the Return phase, the production by the JPN L1 group in the F1 Increase condition returned to baseline; however, production in the F1 Decrease condition did not return to baseline. This asymmetrical de-adaptation pattern during the Return phase was previously observed in a similar study using English native speakers (MacDonald *et al.*, 2011). Although it is unclear why and how these differences occurred, such pattern does not seem to be unique and specific to the language of the target word or speakers' native languages per se.

The current results are relevant to speech production models, such as the DIVA model (for review, see Guenther and Vladusich, 2009). According to the DIVA model, articulation goals are not acoustic points but "convex regions in

orosensory coordinates defining the shape of the vocal tract (Guenther, 1995, p. 595)." The idea that articulatory goals are defined by a region or space is similar to the idea that goals are phonemic categories as the current study proposes.

In summary, the data obtained in the current study indicate that the compensatory response to formant-shifted feedback is not based on purely acoustic processing. Rather, some level of phonological processing appears to influence the behavior. In order to fully examine the role phonological processing plays in compensatory behavior, we will need to examine speakers of other languages and conduct vowel perception measurements.

ACKNOWLEDGMENTS

This research was supported by the National Institute of Deafness and Communicative Disorders Grant No. DC-08092 and the Natural Sciences and Engineering Research Council of Canada. We wish to thank Queen's School of English at Queen's University for assistance in recruiting participants, Bryan Burt for his assistance in data collection, and Jaime Forsythe for her comments on the manuscript.

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.

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.

Cai, S., Ghosh, S. S., Guenther, F. H., and Perkell, J. S. (**2010**). "Adaptive auditory feedback control of the production of formant trajectories in the Mandarin triphthong /iau/ and its pattern of generalization," J. Acoust. Soc. Am. **128**, 2033–2048.

Chen, Y., Robb, M. P., Gilbert, H. R., and Lerman, J. W. (**2001**). "Vowel production by Mandarin speakers of English," Clin. Linguist. Phonetics **6**, 427–440.

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

Guenther, F. H. (**1995**). "Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production," Psychol. Rev. **102**, 594–621.

Guenther, F. H., and Vladusich, T. A. (**2009**). "A neural theory of speech acquisition and production," J. Neurolinguistics, doi:10.1016/j.jneuroling.2009.08.006.

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

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

Kuhl, P. K. (**1991**). "Human adults and human infants show a 'perceptual magnet effect' for prototypes of speech categories, monkeys do not," Percept. Psychophys. **50**, 93–107.

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

MacDonald, E. N., Purcell, D. W., and Munhall, K.G. (**2011**). "Probing the independence of formant control using altered auditory feedback," J. Acoust. Soc. Am. **129**, 955–966.

Manuel, S. Y. (**1990**). "The role of contrast in limiting vowel-to-vowel coarticulation in different languages," J. Acoust. Soc. Am. **88**, 1286–1298.

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

Ng, M. L., Chen, Y., and Sadaka, J. (**2008**). "Vowel features in Turkish accented English," Int. J. Speech-Language Pathology **10**, 404–413.

Nishi, K., and Kewley-Port, D. (**2007**). "Training Japanese listeners to perceive American English vowels: Influence of training sets," J. Speech, Lang. Hear. Res. **50**, 1496–1509.

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

Polka, L., and Bohn, O.-S. (**2003**)."Asymmetries in vowel perception," J. Acoust. Soc. Am. **122**, 1111–1129.

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

Schenk, B. S., Baumgartner, W. D., and Hamzavi, J. S. (**2003**). "Effects of the loss of auditory feedback on segmental parameters of vowels of postlingually deafened speakers," Auris Nasau Larynx **30**, 333–339.

Shiller, D. M., Sato, M., Gracco, V. L., and Baum, S. R. (**2009**). "Perceptual recalibration of speech sounds following speech motor learning," J. Acoust. Soc. Am. **125**, 1103–1113.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., Nishi, K., and Jenkins, J. J. (**1998**). "Perceptual assimilation of American English vowels by Japanese listeners," J. Phonetics **26**, 311–344.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., and Nishi, K. (**2001**). "Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners," J. Acoust. Soc. Am. **109**, 1691–1704.

Stevens, K. N. (**1989**). "On the quantal nature of speech," J. Phonetics, **17**, 3–45.

Vance, T. J. (**1987**). *An Introduction to Japanese Phonology* (State University of New York Press, Albany, NY), p. 11.

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.

Waldstein R. S. (**1990**). "Effects of postlingual deadness on speech production: Implications for the role of auditory feedback," J. Acoust. Soc. Am. **88**, 2099–2114.

Wang, H., and van Heuven, V. J. (**2006**). "Acoustical analysis of English vowels produced by Chinese, Dutch and American speakers," in *Linguistics,* edited by J. M. van de Weijer and B. Los (Benjamins, Amsterdam/Philadelphia), pp. 237–248.