

Language dependent vowel representation in speech production

Takashi Mitsuya^{a)} and Fabienne Samson

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

Lucie Ménard

Département de Linguistique, Université du Québec à Montréal, Pavillon J.-A. De Séve, 320 rue Sainte-Catherine Est, Montréal, QC H2X 1L7 Montréal, Québec, Canada

Kevin G. Munhall

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

(Received 20 December 2012; revised 21 February 2013; accepted 26 February 2013)

The representation of speech goals was explored using an auditory feedback paradigm. When talkers produce vowels the formant structure of which is perturbed in real time, they compensate to preserve the intended goal. When vowel formants are shifted up or down in frequency, participants change the formant frequencies in the opposite direction to the feedback perturbation. In this experiment, the specificity of vowel representation was explored by examining the magnitude of vowel compensation when the second formant frequency of a vowel was perturbed for speakers of two different languages (English and French). Even though the target vowel was the same for both language groups, the pattern of compensation differed. French speakers compensated to smaller perturbations and made larger compensations overall. Moreover, French speakers modified the third formant in their vowels to strengthen the compensation even though the third formant was not perturbed. English speakers did not alter their third formant. Changes in the perceptual goodness ratings by the two groups of participants were consistent with the threshold to initiate vowel compensation in production. These results suggest that vowel goals not only specify the quality of the vowel but also the relationship of the vowel to the vowel space of the spoken language.

© 2013 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4795786>]

PACS number(s): 43.70.Mn, 43.70.Kv, 43.70.Bk [ZZ]

Pages: 2993–3003

I. INTRODUCTION

Producing speech sounds is a process through which phonological representations in the mind are transformed into physical entities—movements of the vocal tract. The perception of speech involves the opposite transformation from the physical (sound waves) to the mental world of phonological and lexical representation. While it is broadly believed that “in all communication, whether linguistic or not, sender and receiver must be bound by a common understanding of what counts: What counts for sender must count for the receiver, else communication does not occur” (Liberman, 1996, p. 31), a detailed account of the relationship between perception and production has been lacking. Empirically, strong relationships between the variance in producing and perceiving speech have been hard to demonstrate (e.g., Newman, 2003). In part, this difficulty is due to vagueness in our understanding of the specific goals of articulation.

Historically, phonological units have been depicted as sparse representations that significantly underspecify what articulation must accomplish (Anderson, 1985). The set of distinctive features, for example, was meant to efficiently

code what was contrastive about phonological units not to completely characterize sounds or serve as a control mechanism for articulation. The strongest evidence of this historical view is that timing, one of the most essential aspects of speech motor control has little or no presence in phonology (Lisker, 1974). Recently, a number of studies have suggested that both perception and production involve very detailed representations of the structure of speech (e.g., Pisoni and Levi, 2007). In other words, the representations required for actually using speech units are far richer than the ones required to create typologies of languages.

In speech production, a traditional way to depict the planning of a sound sequence is to contrast the control of articulators that have specific linguistic goals in the sequence with ones that are unspecified for some sounds and thus free to vary (Henke, 1966). Recent evidence suggests that this is not a true characterization of vocal tract control. When force perturbations are applied laterally to the jaw in a manner that has no acoustic consequences and in a spatial direction that has no relation to phonetic goals, the speech motor system still rapidly compensates to preserve the jaw trajectory (Nasir and Ostry, 2006). These results suggest that everything in the vocal tract has detailed control specifications during speech for reasons such as stability and coordination as well as phonetic goals.

In perception, adults and young infants are sensitive to the fine phonetic detail of utterances. Fine phonetic detail

^{a)}Author to whom correspondence should be addressed. Electronic mail: takashi.mitsuya@queensu.ca

has become a code phrase for reliable information in speech acoustics that is not captured by traditional phonological representations but is clearly used in communication (Nguyen *et al.*, 2009). In adults, people adjust their conversational acoustics in subtle ways to signal meaning, and listeners are remarkably attuned to these fine details (Local, 2003). Infants in the first year of life are already sensitive to this fine phonetic detail and make discriminations based on it (Swingley, 2005). This perceptual sensitivity suggests that talkers are reliably producing much richer categorical information than a distinctive feature assessment would lead one to expect.

In this manuscript, we present new data on the representation of the same vowel quality in French and English. Our approach is to perturb the auditory feedback that talkers hear when they produce a speech token and measure the compensatory behavior.

By systematically studying normal articulatory tendencies and the ways in which the talkers preserve the integrity of their phonological speech goals, we can get a more robust picture of sound representation in production. This can be done with an auditory feedback paradigm where produced acoustic parameters are manipulated in real time, and we observe how speakers compensate for the perturbation.

Although, the importance of the self perception-production relationships has already been demonstrated with post-lingually deafened and hearing impaired individuals (Waldstein, 1990; Cowie and Douglas-Cowie, 1983, 1992; Schenk *et al.*, 2003), laboratory examinations with a real-time auditory feedback paradigm have provided us with more detailed characteristics of how an acoustic difference between an intended productive target and the perturbed feedback error is compensated for with various speech characteristics such as loudness (Bauer *et al.*, 2006), pitch (Burnett *et al.*, 1998; Jones and Munhall, 2000), vowel formant frequency (Houde and Jordan, 2002; Purcell and Munhall, 2006; Villacorta *et al.*, 2007; MacDonald *et al.*, 2010; MacDonald *et al.*, 2011) and fricative acoustics (Shiller *et al.*, 2009).

In vowel perturbation studies, the auditory feedback for the first and second formants (F1, F2, respectively) is increased or decreased in hertz while speakers are producing the vowel in a monosyllabic context (e.g., “head”). Speakers receive real-time feedback of a vowel slightly different from the one intended. In response, they spontaneously change the formant production in the opposite direction of the perturbation applied although the magnitude of compensation is always partial (Houde and Jordan, 2002; Purcell and Munhall, 2006; Villacorta *et al.*, 2007, MacDonald *et al.*, 2010; MacDonald *et al.*, 2011). The consistent patterns of compensatory formant production reported in these experiments have suggested that the productive target is not a specific acoustic point, but rather it is a more broadly defined acoustic region (Guenther, 1995; Guenther *et al.*, 1998), which may reflect a phonemic category (MacDonald *et al.*, 2012).

The International Phonetic Association’s vowel quadrangle assumes that all of the world’s vowels can be slotted into equivalence categories with a small number of dimensions representing the differences between vowels.

However, for the last 50 years or more, phoneticians have reported differences between similar vocal qualities across languages (for a review, see Jackson and McGowan, 2012). While intrinsic F0 vowel patterns are generally consistent across languages (Whalen and Levitt, 1995), similar vowels differ on many dimensions depending on the language’s vowel and consonant inventory (e.g., Bradlow, 1995). This suggests that the goal or phonetic intention of talkers differs depending on the language and thus depending on the language-conditioned phoneme. By studying the specifics of the sensorimotor control of speech sounds across languages, we can characterize in detail the representations that are responsible for the exquisite coordination of articulation.

A recent study by Mitsuya and his colleagues (2011) examined a cross-language difference of formant compensatory production of a vowel that is typologically similar in English and Japanese, and they found a difference across these language groups in terms of the magnitude of compensation. According to the report, when F1 of /ɛ/ (“head”) was decreased in hertz, which made the feedback sound more like /i/ (“hid”), the magnitude of compensation across the language groups was similar; whereas, when F1 was increased in hertz, which made the feedback sound more like /æ/ (“had”), English speakers compensated as much as they did in the decreased condition, but the Japanese speakers did not compensate as much. The same pattern of asymmetry was also observed among another group of Japanese speakers, when they were examined with their native Japanese vowel /ɛ/ which is somewhat similar with the English vowel in the word “head.”

Mitsuya *et al.* (2011) argued that the difference in behavior between languages reflects how the target vowel is represented in the F1/F2 acoustic space and its relation to the nearby vowels. That is, for English speakers, both increased and decreased F1 feedback may have resulted in a comparable perceptual change from the intended vowel, which in turn, elicited a comparable amount of compensation. For Japanese speakers, however, the vowel found in “had” is unstable both perceptually (Strange *et al.*, 1998; Strange *et al.*, 2001) and productively (Lambacher *et al.*, 2005; Mitsuya *et al.*, 2011). Therefore when perturbation sounded more like “had,” the sound might have been heard as an acceptable instance of the intended vowel for “head.” Whereas with the more “hid” like feedback, the vowel might have been perceived to be different because the vowel is known to be perceptually assimilated to the Japanese vowel /i/ (a vowel comparable to the English vowel found in “heed” but much shorter; Strange *et al.*, 1998; Strange *et al.*, 2001). In short, these results indicate that how much speakers compensate may be related to their vowel representation in the F1/F2 acoustic space and the perceptual consequence of perturbation.

The exact mechanism of how compensatory behavior is related to such a representation is still not well understood. What exactly is the phonetic target, and what is required for the error feedback system to detect an error in the acoustics of a sound that has just been spoken? Although Mitsuya *et al.* (2011) suggest that the target reflects the phonemic category of the speech segment, their results might have been unique to (1) compensatory production of F1 and to (2) the

language groups contrasted (English versus Japanese). To further understand the nature and parameters of phonemic processes on compensatory formant production and its mechanism, the current study (1) examined compensatory production of F2 and (2) contrasted different language groups (English versus French).

French and English have different vowel representations and vowel space density. Specifically, the two languages differ in the use of (or the lack of) rounded front vowels. The front vowels /i, e, ε/ in French, found in French words/letter, “i” “é,” “haie,” have a canonical rounded vowel counterpart /y, ø, œ/ (found in “u,” “eux,” “œuf”). Lip rounding as a linguistic feature is binary (rounded or unrounded), but the articulatory postures between these cognates differ in many ways beyond the activity of the orofacial muscle groups associated with the lips (Schwartz *et al.*, 1993). Consequently, there are a host of acoustic differences associated with “lip rounding.” One of such consequences is lowered F2 value. Because of the association between lip rounding and lower F2, with synthesized two-formant vowels, lowering F2 alone can elicit the perception that vowel is slightly rounded for speakers of a language with rounded front vowels (Hose *et al.*, 1983). However, those who do not produce rounded front vowels such as native speakers of English may not be as sensitive to such acoustic changes.

If the F2 of French front vowel /ε/ (English equivalent of the vowel in the word “head”) is lowered, the value will be comparable to that of another French vowel /œ/ in the word “œuf.” Consequently, such a perturbation should induce a perceptual change among French speakers due to their /ε-œ/ proximity in F2. In response, compensatory formant production (raising F2) should be observed. However, for English speakers who do not have the /œ/ phoneme in their vowel inventory, the same magnitude of perturbation of F2 may not result in a comparable degree of perceptual change. The difference in vowel representation and their internal structure across the two language groups should reflect the perception of the feedback (e.g., Kuhl, 1991). In short, the same amount of acoustic perturbation of F2 may not yield a similar perceptual change between the two language groups. Compensatory behavior is expected to reflect the speaker’s internal structure of the vowel system of their language.

In the current study, we measured the perceptual distinctiveness and the sensitivity to such changes by perceptual goodness rating of the vowel being tested and attempted to find a relationship between a goodness rating and formant compensatory production. As past studies have reported, the majority of the speakers did not typically notice that the quality of the vowel changed when perturbation was applied (see Purcell and Munhall, 2006). If speakers tend to hear the produced sound as what was intended, it is more informative to assess goodness of the produced sound as an intended target. For this reason, we decided to employ a goodness rating task instead of a categorization task.

Because perceptual goodness of a vowel is influenced by how the vowel is represented (thus, language specific), and because compensatory formant production is mediated by such perceptual processes, French speakers and English speakers should show different behaviors. Specifically, it

was hypothesized that French speakers would compensate (1) sooner (i.e., with a smaller perturbation) and (2) with a greater maximum compensation magnitude compared to English speakers.

II. EXPERIMENT

A. Participants

Forty-one female speakers took part in this experiment. Nineteen of them were native Quebec French speakers (FRN, hereafter) from the community of Université du Québec à Montréal. The mean age of this group was 27.84 (range: 20–41 yr old). Because Montréal is virtually a bilingual city, we assessed how much English exposure and usage that our participants had using a self-assessing language background survey, The Language Experience and Proficiency Questionnaire (LEAP-Q; Marian *et al.*, 2007). One participant did not complete the survey correctly, thus the following survey results do not include this individual; however, she indicated that she does not speak, read, or communicate in English in her daily activities. Based on her responses, the average proportion of exposure to Quebec French was 82.8% (s.d. = 12.6%), while to English was 15.7% (10.9%). The average proportion of speaking Quebec French was 92.7% (9.8%), whereas that of English was 6.3% (9.5%).

The remaining 22 participants were native English speakers from the Queen’s University community in Ontario Canada. The majority of them had taken French classes but based on the LEAP-Q report, they did not consider themselves bilingual, except for five participants, who indicated that they would be likely to choose both French and English when they were to have a conversation with an English-French bilingual. Because the contrast that we were interested in examining across the language groups was based on the presence or absence of the representation /œ/, we did not want to include English speakers whose linguistic experience may have led to them having this vowel category. For this reason, these five individuals were removed from analyses. The mean age of the remaining 17 English speakers (ENG, hereafter) was 20.81 (range: 18–25 yr old). None of the participants reported speech or language impairments, and all had normal audiometric hearing threshold over a range of 500–4000 Hz.

B. General procedure

Participants first completed a vowel exemplar goodness rating experiment and then participated in the production experiment. The order of these experiments was fixed because we did not want the perturbed feedback during the production experiment to affect the speakers’ goodness/perceptual function of the vowel being tested. ENG received instructions in English, while FRN were given the same instructions in French by a researcher whose native language is Quebec French.

C. Perception experiment

A goodness rating experiment was carried out, in which participants listened to the members of two continua /œf-ɛf/

and /ɛf-if/ over headphones (SONY MDR-XD 200). Both continua were created based on a natural token of “œuf,” “F,” and “if” produced by a French-English bilingual female speaker in such a way that the first four formants were manipulated incrementally between the end points of the natural tokens (“œuf-F” and “F-if”). This method created 11 members for each continuum (thus, 21 unique tokens overall, with the natural token of “F” as one of the end points shared by the two continua). Participants’ task was to rate how good each sound was as “F /ɛf/” using a scale of 1 (poor) to 7 (good) by pressing numeric keys on a computer keyboard. Participants were given the natural token of “F” as the referent sound several times before starting the experiment to ensure that they know what was considered to be a good instance of “F.” During the experiment, each token was randomly presented from the full set of 21 stimuli in the two continua 10 times. The participants were naïve to the fact that two continua were being tested during the perception experiment.

D. Production experiment

1. Equipment

Equipment used in this experiment was the same as that reported in [Munhall et al. \(2009\)](#), [MacDonald et al. \(2010\)](#), [MacDonald et al. \(2011\)](#), and [Mitsuya et al. \(2011\)](#). Participants were tested in a sound attenuated room in front of a computer monitor with a headset microphone (Shure WH20) and headphones (Sennheiser HD 265). The microphone signal was amplified (Tucker-David Technologies MA 3 microphone amplifier), low-pass filtered with a cutoff frequency of 4.5 kHz (Hrohn-Hite 3384 filter), digitized at 10 kHz, and filtered in real-time to produce formant shifts (National Instruments PXI-8106 embedded controller). The manipulated speech signal was then amplified and mixed with speech noise (Madsen Midimate 622 audiometer). This signal was presented through the headphones worn by the speakers. The speech and noise were presented at approximately 80 and 50 dBA SPL, respectively.

2. Estimating model order (Screener)

An iterative Burg algorithm ([Orfanidis, 1988](#)) was used for the online and offline estimation of formant frequencies. The model order, a parameter that determines the number of coefficients used in the auto-regressive analysis, was estimated by collecting 11 French vowels for FRN, and 8 English vowels for ENG, prior to the experimental data collection. The 11 French vowels were /i, y, e, ø, ε, œ, a, α, ɔ, o, u/, seven of which (/i, y, e, ø, ε, o, u/) were presented in a /V/ context, and the visual prompts for those vowels were “i,” “u,” “é,” “eux,” “haie,” “eau,” “ou” (respectively). The vowels /a, α/ were presented in a /pVt/ context (“patte,” “pâte”), and the vowel /ɔ/ was in a /Vt/ context (“hotte”). The vowel /œ/ was presented in a /Vf/ context (“œuf”). Along with the /V/ context, the vowel /ɛ/ was also presented in the /Vf/ context (“F”) because this was the word used during the experiment. The eight English vowels were /i, ɪ, e, ε, æ, ɔ, o, u/, all of which were presented in the /hVd/ context

(“heed,” “hid,” “hayed,” “head,” “had,” “hawed,” “hoed,” “who’d,”). Moreover, the vowels /ɔ, ε/ were also presented in the /Vf/ context (“off,” “F”).

The words were randomly presented on a computer screen in front of the participants, and they were instructed to say the prompted word without gliding the pitch. The visual prompt lasted 2.5 s, and the inter-trial interval was approximately 1.5 s. For each participant, the best model order was selected based on minimum variance in formant frequency over a 25 ms segment in the middle portion of the vowel ([MacDonald et al., 2010](#)). The utterances in this experiment were analyzed with model orders ranging from 8 to 12 with no difference between the language groups.

3. Online formant shifting and detection of voicing

Voicing detection was done using a statistical, amplified-threshold technique, and the real-time formant shifting was done using an IIR filter. The Burg algorithm ([Orfanidis, 1988](#)) was used to estimate formants and this was done every 900 μs. Based on these estimates, filter coefficients were computed 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.

4. Offline formant analysis

Offline formant analysis was done using the same method reported in [Munhall et al. \(2009\)](#) and [MacDonald et al. \(2010\)](#), and [MacDonald et al. \(2011\)](#). An automated process estimated the vowel boundaries in each utterance based on the harmonicity of the power spectrum. These estimates were then manually inspected and corrected if required. A similar algorithm to that used in online shifting was employed to estimate the first three formants from the first 25 ms of a vowel segment. The estimation of formants was repeated with the window of 1 ms until the end of the vowel segment. For each vowel, a steady state value for each formant was calculated based on the average of the estimates from 40% to 80% of the vowel duration. These estimates were then inspected and if any estimates were incorrectly categorized (e.g., F1 being mislabeled as F2, etc), they were corrected by hand.

5. Experimental phases

During the experiment, the participants produced “F” /ɛf/ 120 times, prompted by a visual indicator on the monitor. The 120-trial session was broken down into four experimental phases (Fig. 1). In the first phase, baseline phase (trials 1–20), the participants received normal auditory feedback (i.e., amplified with noise added but no shift in formant frequency). In the second phase, ramp phase (trials 21–70), participants produced utterances while receiving feedback during which F2 was decreased with an increment of 10 Hz at each trial over the course of 50 trials. This made a –500 Hz perturbation for speakers’ F2 at the end of this phase. This phase was followed by the hold phase (trials 71–90) in which the maximum degree of perturbation

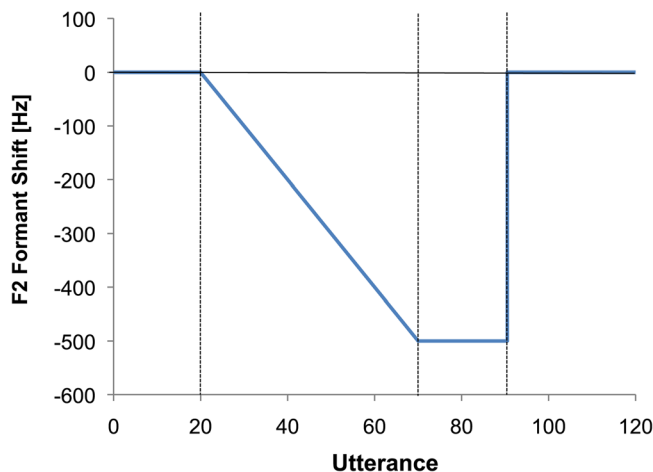


FIG. 1. (Color online) Feedback perturbation applied to the second formant. The vertical dash lines denote the boundaries of the four phases: Baseline, ramp, hold, and return.

(−500 Hz) was held constant. In the final phase, the return phase (trials 91–120), the participants received normal feedback (i.e., the perturbation was removed abruptly at trial 91).

III. RESULTS

A. Perception study

The group averages of /ε/ goodness ratings of the /œf-εf/ and /εf-if/ continua were calculated by averaging 10 goodness ratings for each of the 21 tokens for each individual participant. This experiment was specifically assessing the decline of /ε/ perceptual goodness along the two continua, thus we first analyzed the rating of the natural token of /εf/ (Token 11) across the language groups. On average FRN rated token 11 as 6.71 [standard error (s.e.)=0.09] while ENG as 6.29 (0.11), and the difference was significant ($t[34]=3.04$, $p < 0.05$), although both groups rated token 11 the highest. The natural token of /εf/ was produced by a French-dominant French-English bilingual speaker, thus the produced vowel was slightly assimilated to that of French. We calculated a Euclidian distance in the F1/F2 acoustic space between the formant values of token 11 and the average formant values of /εf/ produced in the /εf/ context for each participant that were collected during the screening session. The average distance from token 11 to the average formant values of /ε(f)/ among FRN was 103.06 (s.e. = 16.80) while for ENG, the distance was 386.37 (68.01), and the difference between groups was significant ($t[34]=11.98$, $p < 0.01$). This might have been why ENG did not rate the natural token /εf/ as high as FRN even though both groups were given token 11 as a referent token several times before the experiment began. Due to the group difference in the rating of the referent token, the degree of goodness decline is not easily compared, therefore, we normalized the rating for each individual so that the rating of token 11 was given the value 7, and the difference between the raw score for token 11 and the score 7 was equally added to all of the average rating responses of the other continuum tokens.

An analysis of variance (ANOVA) with language group as a between-subjects and continuum tokens as a within-subjects

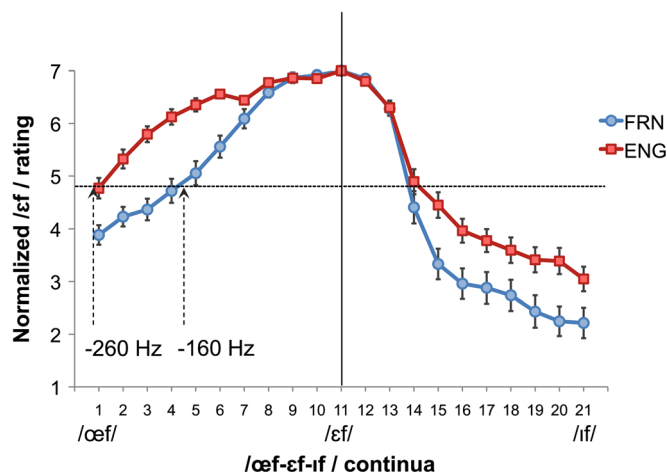


FIG. 2. (Color online) Normalized goodness rating of /εf/ of the continua /œf-εf/ and /εf-if/. The ratings were normalized such that token 11 (naturally produced “F /εf/”), was given a rating of 7. The solid vertical line separates the two continua, and the arrowed vertical lines indicate the magnitude of F2 perturbation needed for speakers to start compensating (i.e., threshold), −160 Hz for French speakers, and −260 Hz for English speakers. The horizontal line represents the goodness score corresponding to the threshold of the language groups. The error bars indicate 1 standard error.

factor was conducted for each of the two continua. For the /œf-εf/ continuum, both main effects were significant (language: $F[1, 34] = 20.55$, $p < 0.05$; tokens: $F[10, 340] = 141.85$, $p < 0.05$), but more importantly, the interaction between the two factors was significant ($F[10, 340] = 14.97$, $p < 0.05$). As can be seen in Fig. 2, while FRN speakers’ goodness ratings robustly decline as tokens are moving toward token 1 from the reference (token 11), the decline of English was more moderate. This confirms that FRN are more sensitive to lowered F2 because perception of the sound moves away from the category of /ε/ and toward the category of /œ/, while ENG are more tolerant of the manipulation.

The results for the /εf-if/ continuum also revealed a significant main effect of language ($F[1,34] = 6.11$, $p < 0.05$) and continuum tokens ($F[10,340] = 256.81$, $p < 0.01$), and a significant interaction of the two factors ($F[10,340] = 5.03$, $p < 0.01$). FRN seems to show more sensitivity to this continuum as well. Due to the difference in density of the vowels in the front mid to close region across the two languages, this perceptual difference was also expected. For English there is no monophthong between /ε-i/ in the F1/F2 acoustic space (note that /e/ is generally diphthongized and much longer in duration, thus qualitatively it is not a monophthong) while in French, there is a monophthong /e/. Because French does not have /i/, the vowel is usually perceptually assimilated to their /i/ (Iverson and Evans, 2007). Thus for Francophones, the /εf-if/ continuum may consist of three vowel categories /ε-e-i/. With the added vowel /e/ in the middle of the continuum, the goodness of /ε/ declined sharply, and this was evident by comparing the slope of the two language groups.

B. Production study

The baseline averages of F1 and F2 values were calculated based on the last 15 utterances of the baseline phase

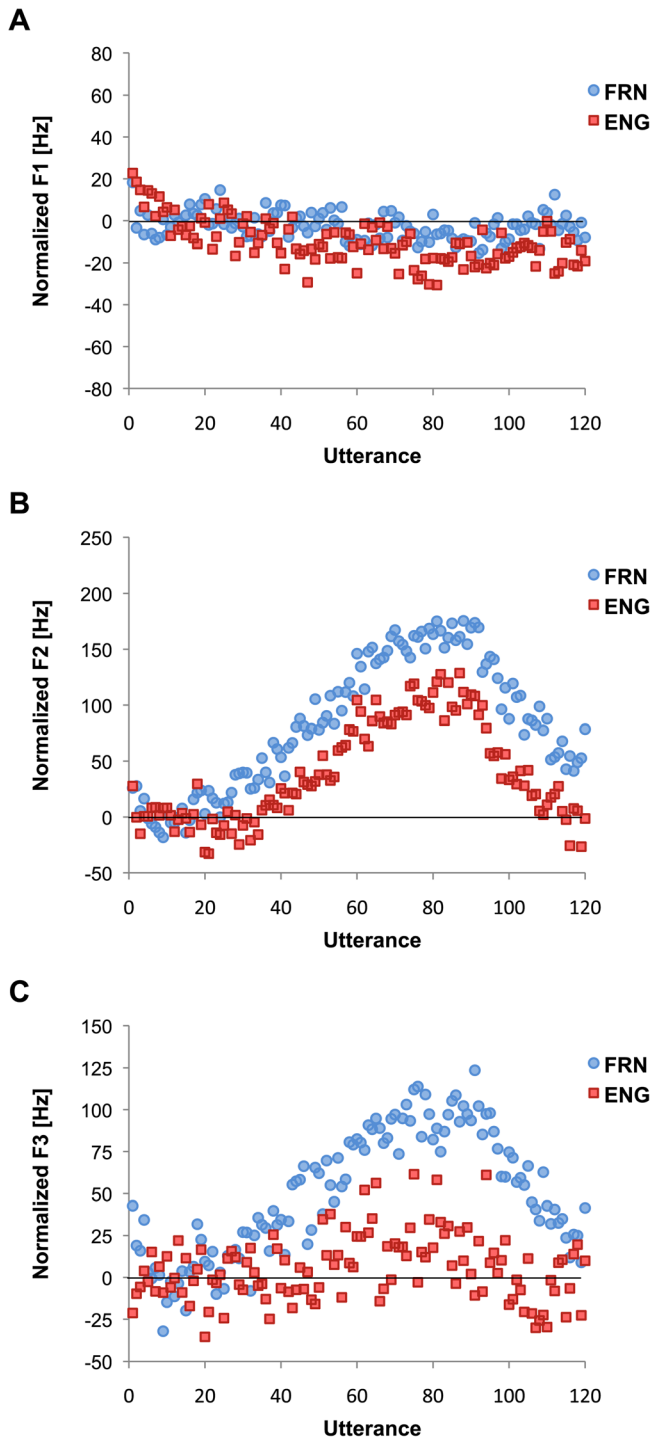


FIG. 3. (Color online) Normalized formant production average across speakers within each language group, (A) first formant, (B) second formant, and (C) third formant.

TABLE I. Average formant values of the last 15 trials of baseline (trials 6–20), hold (76–90), and return (106–120) phases.

	F1			F2			F3		
	Baseline	Hold	Return	Baseline	Hold	Return	Baseline	Hold	Return
FRN	632.29	625.27	631.9	2157.54	2319.43	2223.89	2915.58	3011.95	2949.82
(s.e.)	(10.53)	(11.1)	(11.31)	(23.14)	(25.66)	(23.43)	(29.43)	(29.49)	(29.76)
ENG	864.3	844.7	854.72	1895.87	2002.21	1895.21	3017.56	3037.62	3021.83
(s.e.)	(11.13)	(11.74)	(11.96)	(24.46)	(27.12)	(24.77)	(31.11)	(31.18)	(31.46)

(i.e., utterances 6–20), and the formant values were then normalized by subtracting the participant’s baseline average from each utterance value. The normalized values for each utterance, averaged across speakers, can be seen in Figs. 3(A) (F1) and 3(B) (F2).

To quantify the change in formant production, we measured the average production of each of F1 and F2 by averaging the last 15 utterances of the baseline (utterances 6–20), hold (76–90), and return (106–120) phases (see Table I). For each formant, a repeated measure ANOVA with the experimental phases as a within-subjects and the language groups as a between-subjects factor was conducted to verify the change in production.

For F1, both main effects were significant (phase: $F[2,68] = 6.00$, $p < 0.05$; language: $F[1,34] = 214.54$, $p < 0.001$), and the interaction between the two factors was not significant ($F[2,68] = 1.40$, $p > 0.016$). Overall, ENG had a significantly higher F1 value [854.75 Hz (s.e. = 11.15 Hz)] than FRN [629.82 Hz (10.54 Hz)]. Because the inherent language difference in F1 value is not the main interest of analysis, and because there was no interaction, we collapsed the language groups to see how experiment phases affect speakers’ F1 production. *Post hoc* analysis with Bonferroni adjustment for the experiment phases (α set at 0.016 for three comparisons) revealed that participants produced significantly lower F1 during the hold phase [734.99 Hz (8.08 Hz)] than the baseline phase [748.30 Hz (7.66 Hz); $t[35] = 3.12$, $p < 0.05$], but the difference between the hold and return phases [743.31 Hz (8.23) Hz; $t[35] = -2.33$, $p > 0.016$] and between baseline and return phases ($t[35] = 1.12$, $p > 0.016$) were not significant.

For F2, similarly, both main effects were significant (phase: $F[2,68] = 97.19$, $p < 0.05$; language: $F[1,34] = 83.51$, $p < 0.05$) but more importantly, the interaction of the two factors was also significant ($F[2,68] = 6.39$, $p < 0.05$). To examine the effect of perturbation across language groups, we compared the magnitude of compensation, calculated by averaging the last 15 utterances of the hold phase (utterances 76–90) of the normalized data for each participant. On average, FRN adjusted their F2 by 161.53 Hz (s.e. = 18.75 Hz) for the -500 Hz perturbation, which is approximately 32%, while ENG adjusted their F2 by 106.34 Hz (15.96 Hz), approximately 21%, and this difference was significant ($t[34] = 2.21$, $p < 0.05$).

We also examined when the speakers started changing their F2 production. A change in production was calculated for each speaker, based on the standard error of the productions in the baseline phase. We defined a “change” as three

consecutive productions whose F2 exceeded +3 s.e. from the baseline average. Then after finding such a trial point, the magnitude of perturbation of that trial was calculated. On average, FRN started changing their F2 production with a -162.11 Hz (s.e. = 27.36 Hz) perturbation, while ENG needed a -260.59 Hz (29.56 Hz) perturbation to initiate F2 adjustment, and the difference was significant ($t[34] = 2.45$, $p < 0.05$). Moreover, we ran a bivariate correlation between the threshold and the average maximum of compensation for each language group. We found that neither group showed a significant correlation (both p 's > 0.05), indicating that early compensators did not necessarily compensate more. Interestingly, the perceptual goodness rating of the most /œf/-like tokens was negatively correlated with the average maximum magnitude of compensation among the French speakers ($r[19] = -0.54$, $p < 0.05$) but not with the English speakers ($r[17] = 0.26$, $p > 0.05$). This suggests that it is not likely that sensitivity to detect an acoustic difference from /ε/ alone is related to the amount of compensation, otherwise English speakers should show the same relationship. The asymmetry of these correctional data across the language groups seems to be attributed to the presence/absence of the category /œ/, thus, we speculate that compensation magnitude is related to how the feedback sound is perceived as a member of the neighboring phonemic category.

We also examined the relationship between the difference in F2 values between naturally produced /ε/ and /œ/ and the magnitude of compensation among the French speakers. A bivariate correlation was analyzed and revealed that the coefficient was negative although it was not significant ($r[19] = -0.36$, $p = 0.13$).

Because a lowered F3 is often associated with lip rounding with front vowels (Stevens, 1998; Schwartz *et al.*, 1993), we also examined the production of F3. As can be seen in Table II, the average F3 produced in the unrounded front vowel /ε/ produced in “F” among FRN was 2953.94 Hz (s.e. = 33.07 Hz), while that of ENG was 3013.62 Hz (24.84 Hz), and the group difference was not significant ($t[34] = 1.42$, $p > 0.05$). For FRN, the F3 of the rounded /œ/ produced in “œuf” [2669.75 Hz (27.39 Hz)] was significantly lower than that of the unrounded vowel /ε/ produced in as “F” ($t[18] = 10.99$, $p < 0.001$). Thus for this particular pair of un/rounded vowels, a lower F3 value is associated with rounding.

The F3 production during the experiment was analyzed in the same way as we examined F1 and F2 [Fig. 3(C)]. While ENG did not change the production of F3 in any discernible pattern across the experimental phases, FRN

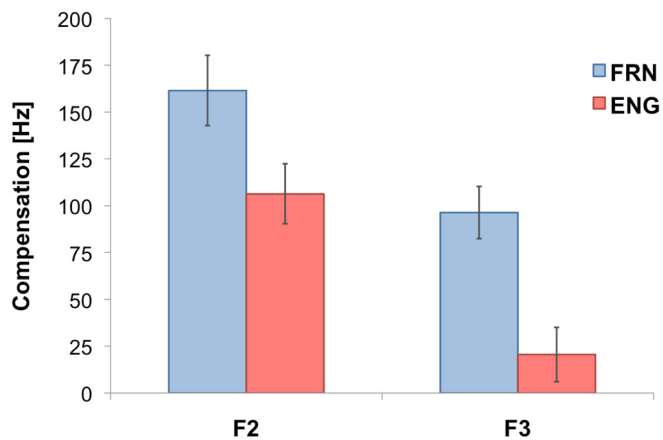


FIG. 4. (Color online) Average compensation in F2 and F3 over the hold phase (i.e., trials 76–90). Compensation is defined as the magnitude of the change in formant frequency from the baseline average. Error bars indicate 1 standard error.

showed a robust change in the F3 production. A repeated measure ANOVA (phase \times language) revealed a significant interaction between the experimental phase and language groups ($F[2.68] = 9.70$, $p < 0.01$). While ENG did not change their F3 production across the experimental phases ($F[2, 32] = 1.36$, $p > 0.05$), FRN significantly increased their F3 production during the hold phase [3011.95 Hz, [(s.e. = 29.49 Hz)] compared to the baseline phase (2915.58 Hz; $t[18] = 6.89$, $p < 0.016$), and to the return phase [2949.82 Hz (29.76 Hz); $t[18] = 6.71$, $p < 0.016$]. Moreover the baseline and the return phases were also significantly different ($t[18] = 3.33$, $p < 0.016$) such that the increased F3 did not fully go back to what it had been in the baseline phase. The magnitude of change in F3 production between the language groups (seen in Fig. 4) was also significantly different ($t[34] = 3.76$, $p < 0.01$), such that FRN increased the F3 value [96. 37 Hz (13.98 Hz)] much more than ENG [20.47 Hz (14.54 Hz)] did.

Because lip rounding is a particularly important articulatory posture for French speakers and because of evidence of co-variance of F2 and F3 for the un/rounded vowels being tested, we examined a correlation between these two formants for each participant and compared the group differences. We calculated Pearson’s correlation coefficient (r) for each of the four experimental phases (note that the first five utterance of the baseline phase were removed, thus the baseline coefficient was based on the last 15 trials; i.e., utterances 6–20) per participant and compared them across the two

TABLE II. Average formant values of /ε/ and /œ/ produced in the /V/, /Vi/ contexts.

	/ε/			/œ/			/œf/		
	F1	F2	F3	F1	F2	F3	F1	F2	F3
FRN	577.20	2396.31	3064.00	624.35	2187.95	2953.94	581.87	1841.62	2669.75
(s.e.)	(11.56)	(28.75)	(31.46)	(9.56)	(27.24)	(33.07)	(10.52)	(17.94)	(27.39)
ENG	742.22	2118.97	3114.94	862.55	1924.64	3013.62			
(s.e.)	(12.11)	(25.80)	(30.30)	(11.80)	(19.21)	(24.84)			

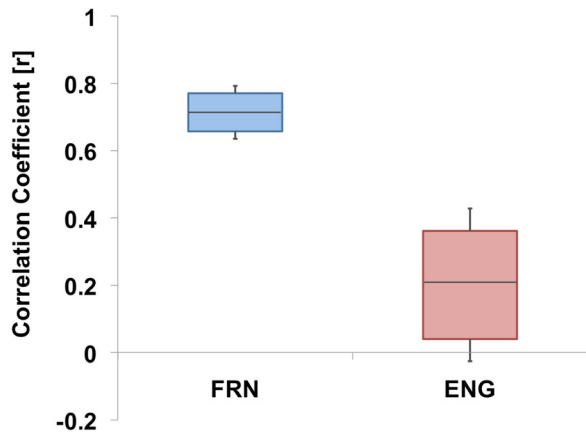


FIG. 5. (Color online) Distribution of correlation coefficients (r) between F2 and F3. The box represents first and third quartiles of r , and the horizontal line in the box indicates the median. The error bars indicate one standard error. The difference in group means was significant ($p < 0.05$).

language groups. A repeated measure ANOVA with experimental phases as a within-subjects and the language group as a between-subjects factor revealed that there was no main effect of experimental phases or the interaction (both F 's < 1 , p 's > 0.05). However, there was a significant effect of language groups ($F[1, 34] = 53.47$, $p < 0.01$), indicating that FRN had a significantly higher positive correlation between F2 and F3 ($X = 0.601$, $s.e. = 0.035$) than that among ENG ($X = 0.231$, $s.e. = 0.037$). Because there was no phase effect, we calculated the overall coefficient per participant across all the experimental trials (except the first five utterances). As can be seen in Fig. 5, all of our French speakers had a significant positive correlation (r values ranging from 0.56 to 0.94, all $p < 0.01$) with the average of 0.71. On the other hand, among the 17 English speakers, only eight speakers had a positive correlation between F2 and F3 (r values ranging from 0.21 to 0.64, all $p < 0.05$) and two speakers having a negative correlation (r 's were -0.21 and -0.31 , both $p < 0.05$). These results clearly indicate that F2 control is strongly associated with F3 for the mid front vowel for French speakers but not for English speakers, but more importantly, it shows that FRN speakers' co-varying F2/F3 production is not to accommodate for the perturbation, rather, the control of these two formants is underlyingly coupled.

In terms of a relationship between the production and the perception data, we compared the formant value at which the groups started compensating to that of the token ratings from the perception experiment. Once again, overall, FRN started compensating approximately with -160 Hz F2 perturbation, which corresponds to utterance 36, while ENG started compensating approximately with -260 Hz, corresponding to utterance 46. The -160 Hz of F2 was comparable to the F2 value midway between tokens 4 and 5 of /œf-ɛf/ continuum (the steady state formant values were taken from the middle 40% of the vocalic part of the tokens), whereas the -260 Hz would have been a token with a slightly lower F2 value than token 1 (thus, somewhat comparable to the natural token of /œf/). On average, the rating of the -160 Hz F2 point among the FRN was slightly lower

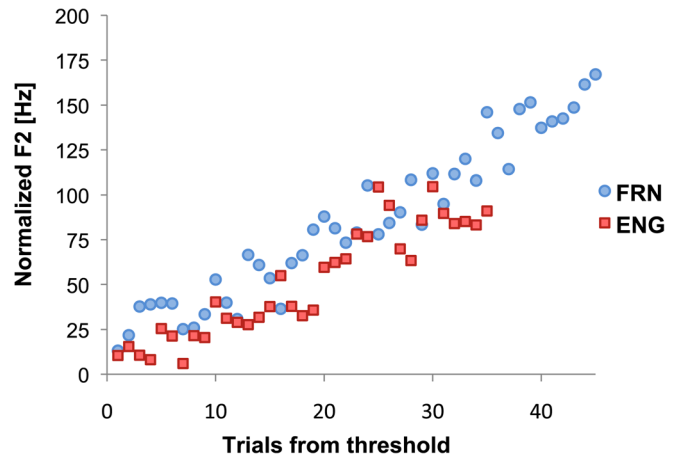


FIG. 6. (Color online) Normalized formant production of F2, aligned at the threshold of compensation (i.e., FRN at trial 36 and ENG at trial 46).

than 5 on the Likert scale. This rating would correspond to that for tokens 1 and slightly below among the ENG (seen in Fig. 2). These perception-production data seem to be comparable across the language groups, such that there seems to be a relationship between the degree of degraded perceptual goodness and the initiation of compensatory production. Moreover if we align the F2 data of the two groups at the threshold, the function of formant compensatory production overlaps almost perfectly. This indicates that although the threshold is defined by language specific phonemic goodness, once the compensation is initiated, the operation of error reduction is similar (Fig. 6) in such a way that the gain is the same across the language groups. These results further confirm that (1) FRN was more perceptually sensitive to the F2 perturbation, (2) a certain decrease in perceptual goodness of the intended phoneme initiates compensation, and (3) once compensation is initiated, the error reduction system seems to behave similarly across the language groups.

One thing to note is that the two language groups produced /ɛ/ slightly differently [see Figs. 7(A) for ENG, 7(B) for FRN; also see Table II; F1: $t[34] = 9.85$, $p < 0.01$; F2: $t[34] = -7.11$, $p < 0.01$]. Moreover, both language groups produced /ɛ/ in the /Vf/ context with a significantly higher F1 (FRN: $t[18] = -8.22$, $p < 0.01$; ENG: $t[16] = -16.27$, $p < 0.01$) and lower F2 (FRN: $t[18] = 14.21$, $p < 0.01$; ENG: $t[16] = 9.32$, $p < 0.01$) compared to the vowel produced in the /V/ context. Thus one can argue that the cross-language difference observed here might have been attributed to the inherent articulatory/acoustic differences across the groups, which were further exaggerated by the phonemic context of the vowel examined. It is not feasible to disentangle the effect of the inherent difference in F1/F2 values across the two language groups versus the perceptual goodness process for the error feedback system, just by looking at the results of F2 magnitude of compensation. However, the results of compensatory production of F3, as well as the threshold of compensation mirroring a specific decrease of goodness rating of the intended vowel strongly suggest that auditory error feedback is specified by phonemic representation of the intended sound.

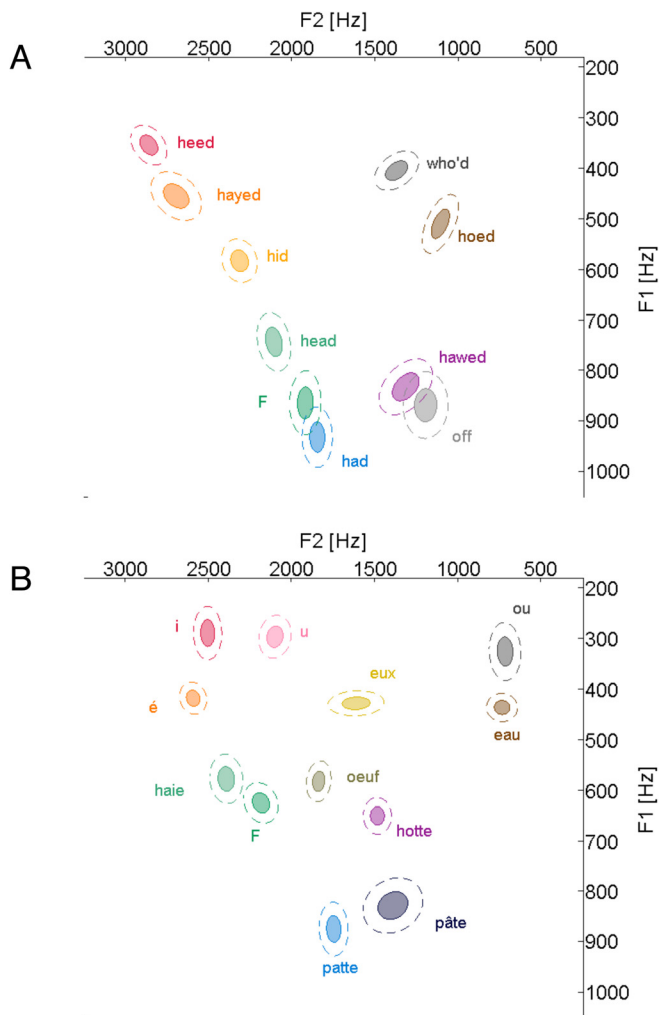


FIG. 7. (Color online) Average vowel space of (A) ENG and (B) FRN in the first formant (F1) and the second formant (F2) acoustic space. The centroid of each ellipse represents the average F1/F2 for that vowel. The solid and dashed ellipses represent 1 and 2 standard deviations, respectively.

IV. DISCUSSION

The current experiment set out to examine the difference in compensatory production for F2 perturbations across two language groups where the same magnitude of perturbation had a different decrease in perceptual goodness of the intended vowel. The results for both languages showed the general pattern of compensatory formant production that has been reported in the literature (Houde and Jordan, 2002; Purcell and Munhall, 2006; Villacorta *et al.*, 2007; MacDonald *et al.*, 2010; MacDonald *et al.*, 2011). More importantly, we saw a cross-language difference between FRN and ENG. French speakers (1) altered their F2 production in response to smaller perturbations, (2) showed greater maximum compensations than ENG did, and (3) showed co-varying F3 compensation with F2, which was not observed among the English speakers. Furthermore, the two language groups initiated compensatory behavior when perceptual goodness of /*ɛ*/ decreased by a similar amount.

These results further confirm that feedback error operation does not involve simply reducing the acoustic error. Instead compensatory behavior is related to how the

feedback is perceived with relation to its nearby vowels, reflecting language-specific phonemic processes. The original hypothesis by Mitsuya *et al.* (2011) stated that the behavior was somehow related to “acceptability” of the perturbed token as the intended sound, which implies processes of phonemic identity and categorization. The current study did not employ a categorical perception task. Therefore we cannot make any conclusion about whether compensation is to maintain perceptual *identity* of the intended token. It is still possible that the threshold of compensation, defined in this study, could have coincided with the categorical boundary between two phonemes. Future studies need to investigate the nature of compensatory threshold more thoroughly.

The importance of phonemic representation in the error reduction system was implicated in a developmental study by MacDonald *et al.* (2012) in which it was found that 2-yr-olds did not compensate for F1/F2 perturbation as adults and older children (4-yr-olds) did. Their results show that the lack of compensatory behavior among the young children was not due to their inherently variable production. MacDonald *et al.* (2012) suggested the possibility that a stable phonemic representation is required for error detection and correction in speech, and sometime between 2 and 4 yr of age such a representation emerges and stabilizes.

The design of the current study, however, does not allow us to tease apart whether the difference in the dimensionality of vowel representation is language general or vowel specific. That is, whether the coupled F2/F3 control is specific to the language in general and thus across all of the vowels among our French speakers, or specific to the particular vowels we examined (i.e., /*œ*/ and /*ɛ*/) is not answered. However, it is testable using a different language group. Unlike French, all of the front vowels of which have a rounded counterpart, Korean has only one front vowel with a rounded counterpart (/*ɛ*/ and /*ø*/; Yang, 1996). Thus we can test whether Korean speakers show co-varying F2/F3 with /*ɛ*/ and compare that with other front vowels. If co-varying F2/F3 compensation is specific to an articulatory/acoustic feature that is phonemically important, then, we would expect that Koreans would show such co-varying production only with /*ɛ*/, but not with other front vowels. With this design, we can thoroughly examine whether or not compensatory formant production is phoneme specific.

The majority of French speakers noticed that the feedback was perturbed in some way, but only two of them reported that they had heard “œuf” while their utterances were perturbed. However, there does not seem to be any relationship between their awareness of the feedback being “œuf” and the magnitude of compensation because these two participants did not compensate significantly more or less than other French speakers (z-scores of these individuals’ magnitude of compensation were -0.6 and -0.5). Similarly, many English speakers noticed that their feedback had been perturbed during the experiment as well, but none reported that it was the vowel quality of the feedback that was being manipulated. Taken together, we can at least rule out the possibility that French speakers’ compensatory production was due to cognitive strategies. Furthermore, a study by Munhall *et al.* (2009) demonstrated that even when speakers were given explicit information about the nature of perturbation and were told not

to compensate for it, they still changed their formant production just as much as naïve speakers. Thus even if there had been a cross-language difference in the level of awareness of the feedback they received, it is questionable that such a difference is the cause of the group difference in magnitude of compensation and F2/F3 co-varying production.

In summary, the current study, along with the findings of Mitsuya *et al.* (2011) strongly demonstrates that the error feedback for formant production is intricately related to speakers' phonemic representation and that this representation contains detailed phonetic information. Thus the hypothesis that error reduction in formant production operates purely to reduce an overall acoustic difference is, once again, rejected. Moreover there is clear evidence that the threshold of compensatory formant production is different across languages. Even though the threshold was different across language groups, the gain function of F2 compensatory production appears to be very similar across the language groups, suggesting a similar error reduction system is in operation. Thus the function of error reduction itself appears to be language universal, while detection of error is language specific. A specific decrease in perceptual goodness of the intended sound within a language initiates compensatory behavior. The data show that the representation of speech goals, at least in vowels, is not a set of articulatory or acoustic features that defines each vowel independently from the other vowels in the inventory. Rather, the target vowel's many phonetic dimensions are intricately represented with those of neighboring vowels. The current study certainly provides evidence for improving the speech error feedback models such as the DIVA model (Guenther, 1995; Guenther *et al.*, 1998), the Neurocomputational model (Kröger *et al.*, 2009) and the State Feedback Control model (Houde and Nagarajan, 2011), so that the models can clearly define an acoustic target for speech production not by the target itself but its relation to other vowels around it and parameterize what the system detects as an error.

ACKNOWLEDGMENTS

This research was supported by National Institute of Deafness and Communicative Disorder Grant No. DC-08092 and the National Sciences and Engineering Research Council of Canada. We would like to thank Paul Plante for assisting with data collection.

- Anderson, S. R. (1985). *Phonology in the Twentieth Century: Theories of Rules and Theories of Representations* (University of Chicago Press, Chicago, IL), pp. 1–350.
- Bauer, J. J., Mittal, J., Larson, C. R., and Hain, T. C. (2006). "Vocal responses to unanticipated perturbations in voice loudness feedback," *J. Acoust. Soc. Am.* **119**, 2363–2371.
- Bradlow, A. R. (1995). "A comparative acoustic study of English and Spanish vowels," *J. Acoust. Soc. Am.* **97**, 1916–1924.
- Burnett, T. A., Freeland, 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.
- Cowie, R. J., and Douglas-Cowie, E. (1983). "Speech production in profound postlingual deafness," in *Hearing Science and Hearing Disorders*, edited by M. E. Lutman and M. P. Haggard (Academic, New York), pp. 183–230.
- Cowie, R. J., and Douglas-Cowie, E. (1992). *Postlingually Acquired Deafness: Speech Deterioration and the Wider Consequences* (Mouton De Gruyter, New York), p. 320.
- 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., Hamoson, M., and Johnson, D. (1998). "A theoretical investigation of reference frames for the planning of speech movements," *Psychol. Rev.* **105**, 611–633.
- Henke, W. E. (1966). "Dynamic articulatory model of speech production using computer simulation," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.
- Hose, B., Langner, G., and Scheich, H. (1983). "Linear phoneme boundaries for German synthetic two-formant vowels," *Hear. Res.* **9**, 13–25.
- Houde, J. F., and Jordan, M. I. (2002). "Sensorimotor adaptation of speech. I: Compensation and adaptation," *J. Speech Lang. Hear. Res.* **45**, 295–310.
- Houde, J. F., and Nagarajan, S. S. (2011). "Speech production as state feedback control," *Front. Hum. Neurosci.* **5**(82), 1–14.
- Iverson, P., and Evans, B. G. (2007). "Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration," *J. Acoust. Soc. Am.* **122**, 2842–2854.
- Jackson, M. T.-T., and McGowan, R. S. (2012). "A study of high front vowels with articulatory data and acoustic simulations," *J. Acoust. Soc. Am.* **131**, 3017–3035.
- Jones, J. A., and Munhall, K. G. (2000). "Perceptual calibration of F0 production: Evidence from feedback perturbation," *J. Acoust. Soc. Am.* **108**, 1246–1251.
- Kröger, B. J., Kannampuzha, J., and Neuschaefer-Rube, C. (2009). "Towards a neurocomputational model of speech production and perception," *Speech Commun.* **51**, 793–809.
- Kuhl, P. K. (1991). "Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not," *Percept. Psychophys.* **50**, 93–107.
- Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., and Molholt, G. (2005). "The effects of identification training on the identification and production of American English vowels by native speakers of Japanese," *Appl. Psycholinguist.* **26**, 227–247.
- Lieberman, A. M. (1996). *Speech: A Special Code* (MIT Press, Cambridge, MA), p. 31.
- Lisker, L. (1974). "On time and timing in speech," in *Current Trends in Linguistics*, edited by T. A. Sebeok (Mouton, The Hague), Vol. 12, pp. 2378–2418.
- Local, J. (2003). "Variable domains and variable relevance: Interpreting phonetic exponents," *J. Phonetics* **31**, 321–339.
- MacDonald, E. N., Goldberg, R., and Munhall, K. G. (2010). "Compensation in response to real-time formant perturbations of different magnitude," *J. Acoust. Soc. Am.* **127**, 1059–1068.
- MacDonald, E. N., Johnson, E. K., Forsythe, J., and Munhall, K. G. (2012). "Children's development of self-regulation in speech production," *Curr. Biol.* **22**, 113–117.
- 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.
- Marian, V., Blumenfeld, H. K., and Kaushanskaya, M. (2007). "The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals," *J. Speech Lang. Hear. Res.* **50**, 940–967.
- Mitsuya, T., MacDonald, E. N., Purcell, D. W., and Munhall, K. G. (2011). "A cross-language study of compensation in response to real-time formant perturbation," *J. Acoust. Soc. Am.* **130**, 2978–2986.
- 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.
- Nasir, S. M., and Ostry, D. J. (2006). "Somatosensory precision in speech production," *Curr. Biol.* **16**, 1918–1923.
- Newman, R. S. (2003). "Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report," *J. Acoust. Soc. Am.* **113**, 2850–2860.
- Nguyen, N., Wauquier, S., and Tuller, B. (2009). "The dynamical approach to speech perception: From fine phonetic detail to abstract phonological categories," in *Approaches to Phonological Complexity*, edited by F. Pellegrino, E. Marsico, I. Chitoran, and C. Coupé (Mouton de Gruyter, Berlin), pp. 5–31.
- Orfanidis, S. J. (1988). *Optimum Signal Processing: An Introduction* (McGraw-Hill, New York), p. 590.

- Pisoni, D. B., and Levi, S. V. (2007). "Some observations on representations and representational specificity in speech perception and spoken word recognition," in *The Oxford Handbook of Psycholinguistics*, edited by M. G. Gaskell (Oxford University Press, Oxford, UK), pp. 3–18.
- Purcell, D. W., and Munhall, K. G. (2006). "Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation," *J. Acoust. Soc. Am.* **120**, 966–977.
- 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.
- Schwartz, J.-L., Beaufemps, D., Abry, C., and Escudier, P. (1993). "Inter-individual and cross-linguistic strategies for the production of the [i] vs [y] contrast," *J. Phonetics* **21**, 411–425.
- 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.
- Stevens, K. N. (1998). *Acoustic Phonetics* (MIT Press, Cambridge, MA), pp. 257–322.
- 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.* **104**, 1691–1704.
- 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.
- Swingle, D. (2005). "11-month-olds' knowledge of how familiar words sound," *Dev. Sci.* **8**, 432–443.
- 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.
- Whalen, D. H., and Levitt, A. G. (1995). "The universality of intrinsic F0 of vowels," *J. Phonetics* **23**, 349–366.
- Yang, B. (1996). "A comparative study of American English and Korean vowels produced by male and female speakers," *J. Phonetics* **24**, 245–262.