SELF-CORRECTION IN L1 AND L2 VOWEL PRODUCTION

Sarah Bakst and Caroline A. Niziolek

University of Wisconsin–Madison
sbakst@wisc.edu, cniziolek@wisc.edu

ABSTRACT

We listen to ourselves while talking, comparing our acoustic output to an internal auditory representation of speech targets. Previous work has shown that speakers are sensitive to their own natural acoustic variability in their native language, steering deviant productions towards auditory targets while speaking [10]. This corrective behavior is evident in the magnitude and direction of vowel formant trajectories over the course of an utterance.

In a language learned in adulthood, auditory targets may be weaker, resulting in less successful self-correction in novel vowel categories. In the current study, participants were recorded producing monosyllabic words in L1 (English) and L2 (French). Speakers' L2 productions showed increased acoustic variability and reduced corrective behavior compared with L1. These results are consistent with weakened auditory representations of speech targets in L2, which may impair the ability to correct one's own productions on-line.

Keywords: L2 phonetics, vowel categories, self-perception.

1. INTRODUCTION

We listen to ourselves while we are talking to ensure that our speech matches what we intended to say and how we intended to say it. Healthy native speakers maintain acoustic consistency by accessing their *auditory feedback*, which they use to detect errors and update motor plans while talking. In a second language, producing native-like phonetic categories is more difficult: speakers have less experience with producing and perceiving these phonemes, so L2 speech motor programs are less well-practiced and internal auditory representations in L2 are weaker. Here, we investigated whether native English speakers who learned French in adulthood are able to detect and correct for errors in these two languages.

Native speakers' sensitivity to their auditory feedback has been consistently shown by studies that alter that feedback in real time. In these studies, speakers compensate for unpredictable experimentally-imposed alterations to their vocal feedback, rapidly adjusting their amplitude [1], pitch [3], or formant frequencies [12, 4, 8] to partially oppose the perceived mismatch between intended and observed speech acoustics. A related line of evidence comes from studies examining on-line responses to natural acoustic variability in produced speech. In these experiments, auditory feedback is left unaltered, but some productions will naturally fall farther from a speaker's acoustic prototype for a given vowel. Self-correction of these productions is evident in how speakers change the course of their formant trajectories while speaking; between the start and middle of a vowel, speakers are able to steer more deviant productions closer to more prototypical productions, reducing variability [11]. Additionally, auditory sensitivity to these more deviant productions is evident in speakers' cortical responses to self-produced speech, which are greater when productions are more acoustically deviant [10].

To achieve this sensitivity, speakers must have a stable sensory representation of the target, which is the basis of comparison on which an error calculation can be performed. Similarly, this target is necessary for speakers to be able to correct themselves accurately. In a second language, unfamiliar phonetic categories have a weaker auditory representation than native categories, impeding the detection of acoustics that deviate from the target representation. Speakers additionally have less practice in implementing motor plans for L2 categories, and may be less able to adjust their productions on-line to produce the intended output.

In this study, we examine variability and self-correction abilities of L1 English speakers producing French as a second language. Where English distinguishes vowels only by height and backness (geographical region of primary lingual constriction), French additionally contrasts whether the lips are rounded or not. English does contrast rounding, but not as an independent feature: in English, only back vowels (e.g. [u]) are ever rounded, and front vowels (e.g. [i]) are always unrounded. In addition to front-unrounded and back-rounded vowels, French additionally has *front-rounded* vowels, such as [œ]. We expect speakers' perceptual targets to be less well-formed for these novel categories. Here we examine formant trajectories during the production of the familiar vowel [i], the French front unrounded mid vowel (similar to English [e] or [ɛ], and denoted here as [ɛ]), and the French low front round vowel [œ].

We hypothesize that L2 vowels will be produced with more variability [6]. Further, because self-correction requires the ability to calculate acoustic distance from a speech target, we hypothesized that there would be less self-correction in French (L2) than in English (L1). We additionally hypothesized that these L2 effects would be greatest in [œ], the vowel which is the most unfamiliar and therefore likely to have the weakest target, and smallest in [i], which is shared across languages, and which occurred in identical phonetic environments in our stimuli.

2. METHODS

2.1. Participants

Nine participants (seven female) took part in the experiment. All procedures were approved by the Institutional Review Board at the University of Wisconsin–Madison and took place at the Medical College of Wisconsin or the Waisman Center at UW–Madison. Four participants were recruited from the Madison and Milwaukee areas, all with at least intermediate-level French experience from high school or university. Pilot data from the co-authors is also included; one is a trained phonetician, and the other has over a decade of experience with French. Both were deemed familiar enough with the tested vowel categories to include their data in the analysis; their data were within the range of data from the seven naive subjects.

2.2. Procedure

Participants were seated in front of a screen and instructed to produce words as they appeared. Each block consisted of 90 speaking trials, and trials were randomized within a block. Blocks alternated between English and French. There were a total of ten blocks (five in each language) for a total of 900 trials (150 per stimulus). Three subjects participated in a concurrent magnetoencephalography (MEG) neuroimaging study; some completed fewer than 900 trials (720, 810, and two with 540 trials) owing to time constraints.

2.3. Stimuli

Three words from each language were produced by participants and are shown in Table 1 below. Stimuli were chosen to minimize effects of coarticulation (no onsets; codas limited to labiodental fricatives where possible). MEG studies require a large number of trials to differentiate signals from noise; for this reason, we used a small number of stimuli with a large number of repetitions.

Table 1: Word production stimuli

English	*Eve* [iv]	*eff* [ɛf]	*add* [æd]
French	*Yves* [iv]	*hais* [ɛ]	*oeuf* [œf]

2.4. Acoustic analysis

For each spoken trial, vowel onset and offset were manually labeled. Within this interval, the first and second formant frequencies (F1 and F2) in Hz were tracked every 5 ms using Praat [2] via the wave_viewer analysis package [7] for Matlab. Formant values were converted to the mel perceptual scale for all analyses.

First, F1 and F2 were averaged across the first 50 ms of each utterance. These values defined the coordinates of the utterance's starting point in an F1/F2 coordinate plane. For each speaker, the center for each vowel category was defined as the median F1 and F2 for that vowel in this early time window. Initial variability for each trial was defined as its distance from the vowel center. Trials were then categorized based on this initial variability. The third of trials with the smallest initial variability, whose starting points were closest to the center, were defined as "center" trials. The third with the greatest initial variability were defined as "peripheral" trials.

Second, F1 and F2 were averaged across a window containing the middle 50% of the vowel for each utterance; these values defined the coordinates for the midpoint of the utterance. The center for each vowel category was similarly redefined as the median F1 and F2 in this middle time window. Mid-trial variability was defined as each trial's distance from this vowel center.

Self-correction was defined as the change in the variability between the beginning and middle of an utterance. For each utterance, mid-trial variability was subtracted from initial variability; self-corrective behavior refers to trials where variability decreased from the first to the second time window. Further details can be found in [9].

3. RESULTS

3.1. Variability

We ran separate ANOVA models to predict variability in the first (initial 50 ms) and second (middle 50%) time windows, with language and vowel as fixed factors and subject as a random factor. Because variability is defined as a distance from vowel center, values approximate only the positive half of a normal distribution (i.e., there are no negative distances); for this reason, we used log-transformed variability as a dependent measure. In the first time window, variability was significantly predicted by language ($p < 0.001$): there was more initial variability in French than in English. Post hoc Tukey testing on the factors of language and vowel revealed that in both languages, [i] had the least variability. In English, there was no difference in variability between [ɛ] and [æ], but in French, all three vowels were significantly different from each other, with the least variability in [i] and the most in [œ]. Finally, [i] and [ɛ] had significantly less variability in English than in French, despite the existence of each phonemic category in both languages. We had hypothesized that speakers would map these categories from L2 directly onto matching L1 categories, but this evidence suggests that speakers in fact have separate categories for these vowels in L1 and L2.

The cross-linguistic patterns observed between similar phonemic categories in English and French during the first time window remained in the second time window. Relationships among English vowels remained the same, as did the relationships among French vowels.

3.2. Self-correction

Self-correction varied between participants and vowels. Group data showing self-correction in all peripheral trials is shown in Figure 1. A mixed-effects ANOVA (same factors as for variability, above) predicting amount of self-corrective behavior in peripheral trials showed main effects of language ($p = 0.006$) and vowel ($p = 0.005$). Post hoc tests revealed more self-correction in English than French.

Tukey tests were used to investigate marginal means differences in language and vowel quality. English [æ] had significantly more self correction than French [i] as well as [ɛ] in both languages. French [ɛ] additionally had significantly less self-correction than English [i]. French vowels did not significantly differ in correction among themselves. Self-correction in peripheral trials was not significantly different between other vowel pairs.

In a separate analysis considering central and middle trials in addition to peripheral trials, language and vowel were both significant ($p < 0.0001$) predictors of self-corrective behavior; there was again more self-correction in English than in French. Post hoc Tukey tests showed no significant differences between English vowels. However, within French, [œ] had significantly more self-correction than all other French vowels and was not different from self-correction in English.

Initial variability (log values) had a significant effect ($p < 0.0001$) when added to the model predicting the self-correction of peripheral trials. In this model, the effect of variability had the greatest coefficient of any factor.

In peripheral trials, log initial variability was positively correlated with self-correction ($r = 0.357, p < 0.0001$). This relationship was stronger in English ($r = 0.45$) than French ($r = 0.32$, both $p < 0.0001$). Thus variability is a significant predictor of self-correction, but its effect is mediated by language.

Finally, to test whether our results could be explained by regression to the mean, we calculated reverse-time changes in variability, redefining the peripheral trials by the second time window and measuring correction in the first time window. An ANOVA comparing self-correction with this time-reversed correction found a significant effect of the additional factor of time direction ($p = 0.01$), with more self-correction for forward than reverse time.

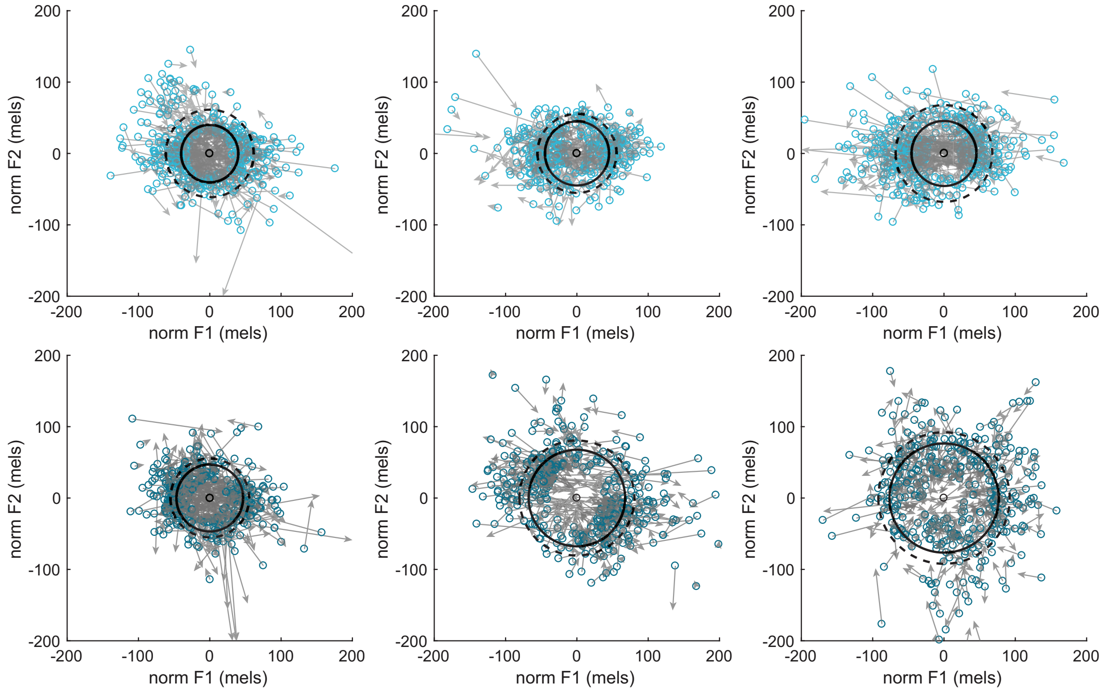
4. DISCUSSION

Subjects showed more variability and less self-correction while speaking L2 than L1, supporting our hypothesis that language inexperience results in a weaker auditory target, hindering the error calculation required to devise a corrective motor plan.

We also found a significant correlation between initial variability and self-correction, suggesting that the detection of this variability may underlie the correction. This correlation may explain why [œ], which should have been the least familiar vowel, did not have significantly less correction than any other vowel. Interestingly, initial variability was significantly greater for French vowels, but it was accompanied by a smaller, not a greater, self-correction; that is, the resulting behavior was not as effective. These results suggest that vowel-specific familiarity is not the only force driving language differences in self-correction.

Vowel familiarity seemed to have some effect on variability. In both time windows, the most unfa-

Figure 1: Self-correction for all subjects (overlaid), English vowels (top row; from left: [i], middle: [ɛ], right: [æ]). French vowels (bottom row; from left: [i], middle: [ɛ], right: [œ]). Open circles denote formants at trial onset and arrowheads denote formants at trial midpoint.



miliar vowel ([œ]) had the most variability, as predicted. For familiar vowels ([i,ɛ]), there was less variability in English than in French both at initial and mid trial. The differences in variability in [i] across languages were unexpected and suggest that variability, even within a single phonemic category, is language-specific. Although [i] and [ɛ] are similar in both languages, speakers have more experience with those categories occurring in English than in French, which may preferentially strengthen their English targets. In fact, there is evidence that speakers had language-specific targets for [i] and [ɛ]: t-tests (Bonferroni-corrected to $\alpha = 0.01$) showed that the mean F1 and F2 were significantly different across languages, with English [i] lower and fronter than French [i], and English [ɛ] lower and backer than French [ɛ].

In summary, vowel targets, variability, and self-correction vary as a function of language experience.

5. CONCLUSION

In order to self-correct, speakers must be able to calculate both the distance they have strayed from their target and a motor plan that will allow them to reach it. These calculations are less practiced when speaking a second language compared to the native language. Speakers may have less sensitivity to fine-grained, within-category differences in unfamiliar phonetic categories, which may impair the ability to detect and calculate errors. Less experience with attaining the target may impair the ability to calculate a motor plan on-line to reach that target.

This study found that successful self-correction is reduced when speaking a second language. While the novel category [œ] underwent a comparable magnitude of self-correction, the final variability was still greater than in native vowel categories. Further, self-correction differed within similar vowel categories across languages, suggesting that even if targets from a new language are mapped onto the native language [5], the production system does not treat the targets as identical to each other.

6. REFERENCES

[1] Bauer, J. J., Mittal, J., Larson, C. R., Hain, T. C. 2006. Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude. *The Journal of the Acoustical Society of America* 119(4), 2363–2371.

[2] Boersma, P., Weenink, D. Praat: doing phonetics by computer [computer program].

[3] Burnett, T. A., Freedland, M. B., Larson, C. R.

1998. Voice F0 responses to manipulations in pitch feedback. *Journal of the Acoustical Society of America* 103, 3153–3161.

[4] Cai, S., Ghosh, S. S., Guenther, F. H., Perkell, J. S. 2011. Focal manipulations of formant trajectories reveal a role of auditory feedback in the online control of both within-syllable and between-syllable speech timing. *The Journal of Neuroscience* 31(45), 16483–16490.

[5] Flege, J. E. 1987. The production of “new” and “similar” phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics* 15, 47–65.

[6] Kartushina, N., Frauenfelder, U. H. 2014. On the effects of l2 perception and of individual differences in l1 production on l2 pronunciation. *Frontiers in Psychology* 5(1246), 1–17.

[7] Niziolek, C. Jan 2015. wave_viewer: First release.

[8] Niziolek, C. A., Guenther, F. H. 2013. Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *The Journal of Neuroscience* 33(29), 12090–12098.

[9] Niziolek, C. A., Kiran, S. 2018. Assessing speech correction abilities with acoustic analyses: Evidence of preserved online correction in persons with aphasia. *International Journal of Speech-Language Pathology* DOI: 10.1080/17549507.2018.1498920.

[10] Niziolek, C. A., Nagarajan, S. S., Houde, J. F. 2013. What does the motor efference copy represent? Evidence from speech production. *The Journal of Neuroscience* 33(41), 16110–16116.

[11] Niziolek, C. A., Nagarajan, S. S., Houde, J. F. 2015. The contribution of auditory feedback to corrective movements in vowel formant trajectories. for ICPhS 2015, T. S. C., (ed), *Proceedings of the 18th International Congress of Phonetic Sciences* number Paper number 1004 Glasgow, UK. Retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS1004.pdf. University of Glasgow.

[12] Tourville, J. A., Reilly, K. J., Guenther, F. H. 2008. Neural mechanisms underlying auditory feedback control of speech. *NeuroImage* 39(3), 1429–1443.