ASSESSING THE EFFECTS OF AUDITORY FEEDBACK PERTURBATION ON PERCEPTION OF SIMILAR VOWELS IN A DIFFERENT LANGUAGE

Madeleine Rees[1], Abbie Bradshaw[2], Matthew H. Davis[2], Brechtje Post[1]

[1]Dept. of Theoretical & Applied Linguistics, University of Cambridge [2]MRC Cognition and Brain Sciences Unit, University of Cambridge
mljr2@cam.ac.uk

ABSTRACT

Auditory feedback perturbation of vowel formants in real time causes the speaker to perceive a change in their own voice, and to change their production to offset the perceived error. This response, known as adaptation, has been shown to influence the perceived category boundary of the perturbed vowel for monolingual English speakers. Here, we added a bilingual dimension by applying formant perturbation to tokens of "head" and "dedo" produced by 23 English-Spanish bilinguals, which moved the intended vowel /ɛ/ or /e/ closer to /i/. The feedback perturbation task was preceded and followed by a forced choice vowel identification task in the language not used in the production task. Results indicate significant adaptation responses to formant perturbation in English dominant speakers, producing English words. However, we observed null effects of auditory feedback perturbation on categorical perception of similar vowels in the other language. The roles of vowel inventory and experimental design are discussed.

Keywords: speech perception, speech production, feedback perturbation, bilingualism

1. INTRODUCTION

The relationship between speech perception and speech production is unclear and highly-debated. One paradigm which can uncover functional links between production and perception is auditory feedback perturbation (AFP). AFP is an experimental technique in which a computer modifies the acoustic cues of produced speech in near-real-time (typically less than 50ms delay). The altered feedback causes the speaker to perceive a change in their own voice and adjust their speech to offset the effects of the perturbation, a process termed adaptation [5]. For example, a speaker instructed to say "head" and given a downshift perturbation in F1 would hear themselves producing an aberrant vowel with a lowered F1 (i.e. "hid"). Over repeated trials, the adaptation response would consist of speakers producing a sound with a higher F1, closer to "had", opposing the perturbation.

The adaptation response is predicted by models like DIVA [16] and SFC [6], in which a difference between expected speech acoustic cues (in accordance with an auditory target for speech production) and observed speech acoustic cues leads to corrective motor commands, in turn generating the adaptation response.

Feedback-induced changes in production of vowel formants have been found in turn to influence vowel category boundaries in a vowel perception task [9]. Adaptation to a downshift of F1 of the /ɛ/ vowel (in the word "head") was associated with an increase in hearing "had" rather than "head" in a follow-up 2AFC vowel identification task. This finding suggests that adaptation to the altered feedback is associated with a reconfiguration of the perceptual boundaries of the perturbed vowel and surrounding vowels. By extension, the change in production induced by AFP appears to influence phonemic representations and auditory targets of the perturbed and nearby vowels.

Production and perception tasks in [9] were carried out in English. So far, no study has performed this experiment across languages with bilingual speakers. Testing bilingual transfer of adaptation allows for further exploratory investigation of whether the adaptation response modifies the perceptual boundary of solely the perturbed vowel, or reconfigures the vowels in the surrounding vowel space more generally and across linguistic dimensions. Nearby vowels in the other language with similar acoustic characteristics may be susceptible to the effects of adaptation on categorical perception if they are also reconfigured during speech motor learning. However, other research indicates that speech motor learning is 'local' [17]. In this study, speech motor learning leading to changes in production of the /iæ/ diphthong did not generalise to different words. This result suggests that the effects of adaptation are highly specific and may not extend to affect similarly close vowels in another language. However, the study in [17] was a jaw perturbation study rather than an AFP study, so the two paradigms are not directly comparable. More information about the reconfiguration of phonemic representations during AFP is a useful addition to speech production models like DIVA and SFC.

The vowel inventories of English and Spanish contain a similar but not identical mid/mid-high front vowel (/ɛ/ and /e/ respectively). Therefore, an appropriate speaker group in which to compare cross-linguistic perceptual boundary change is English-Spanish bilinguals. Here, we report a null effect of feedback perturbation of English /ɛ/ on categorical perception of the /e/-/a/ continuum in Spanish, and of perturbation of Spanish /e/ on perception of the /ɛ/-/æ/ continuum in English.

2. METHODS

2.1. Participants

Participants were 23 English-Spanish bilinguals (12 (Southern British) English L1, 11 Spanish L1, aged 19-47). All of the Spanish L1 participants spoke Peninsular Spanish, except for one Mexican and one Chilean Spanish-speaking participant. Language dominance was calculated using the Bilingual Language Profile [1]: henceforth *language-dominant* will be used to refer to each speaker group. Ethical approval was granted by the departmental ethics committee. Participants were compensated for their time.

2.2. Production experiment design

The experiment took place in the soundproof booth of the Phonetics Lab of the University of Cambridge. Feedback perturbation was configured in MATLAB (MathWorks) using Audapter [3] to introduce a consistent perturbation of -130 mel in F1 and +130 mel in F2.

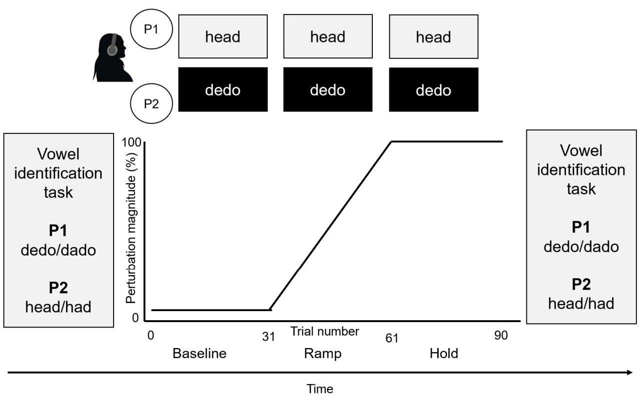

Figure 1: Experimental design showing task order over time and perturbation magnitude over each phase of the production test. Session 1 (noted as P1) was performed first (perception-production-perception) followed by session 2 (P2). The language of the feedback perturbation task was the opposite of the language of the vowel identification task (i.e. task language was switched twice in each session).As participants produced speech, they heard their own voice presented binaurally through headphones in real time. The participant's own voice was mixed with 60dB masking noise to reduce bone conduction.

The English-Spanish task was conducted first. After a pre-test assessing categorical vowel perception in Spanish, the experiment started with a baseline phase in which participants produced 30 "head" utterances with no formant perturbation. Subsequently, a ramp phase introduced the perturbation gradually over 30 trials until the maximum perturbation was reached. Then, in the Hold phase, the maximum perturbation was held constant over 30 trials. At this point, the participant undertook a categorical perception test in Spanish. The Spanish-English task followed the same protocol, with switched language order (see Fig. 1) and with the production stimuli being "dedo". For 16 participants, both English-to-Spanish and Spanish-to-English task runs were conducted in the same session, with a 10–15-minute break between tasks to wash out the effects of the first perturbation. For 6 participants, the two task runs were conducted in separate sessions typically a week apart, due to logistical constraints. Instructions were given in the perturbed language.

Speech was recorded with a Sennheiser cardioid microphone approximately 10cm from the speaker. The audio signal was processed with a SoundDevices MixPre6 integrated soundcard and mixer connected to an ASUS Zenbook Pro laptop via USB-C. The average latency of the Audapter setup was between 16-26ms, with <50ms typically considered acceptable latency for feedback perturbation experiments [15].

2.3. Categorical perception tests

Vowel stimuli in English and Spanish were taken from recordings of "head", "had", "dedo" *(finger)* and "dado" *(dice)* produced by a male and female native speaker of British English and a male and female native speaker of Peninsular Spanish. Two continua of 10 resynthesised vowel tokens between "head" and "had" and "dedo" and "dado", with identical durations, were generated using Tandem-STRAIGHT morphing [7]. Morphing percentages ranged from 0% to 100% in equidistant 11.1% intervals.

Vowel perception tests were run in MATLAB using PsychToolBox-3 [2]. The 10 vowel stimuli were presented 10 times each binaurally through headphones with an ITI of 1s. The gender of the stimuli speaker was matched to the gender of the participant, to ensure that the perception stimuli sounded as similar as reasonably possible to the participant's own production stimuli. Responses were assessed by keyboard press.

2.4. Production data analysis

Adaptation response magnitude was assessed by projecting the vector of each token in the Hold phase onto the vector directly opposing the perturbation (i.e. the vector of perfect compensation).

2.5. Perception data analysis

Categorical perceptual boundary (CPB) midpoints (i.e., the point of subjective perceptual equality) between vowels were calculated in R [14]. Logistic functions were fitted to perception test response data. CPB midpoint was calculated as the number of the vowel token (between 1-10) at which the proportion of responses was 0.5 (i.e. the point of subjective equality).

3. RESULTS

3.1. Feedback perturbation task results

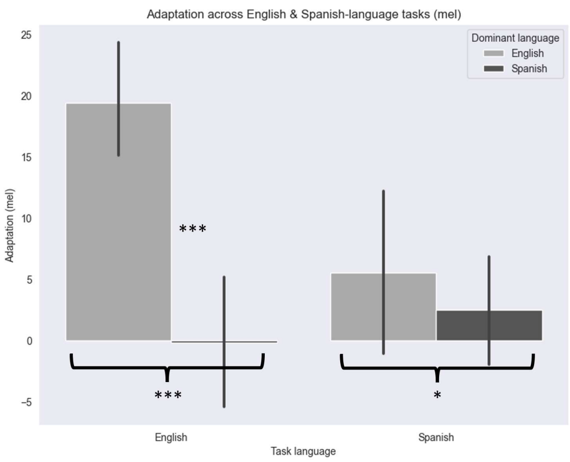

Figure 2: Adaptation responses to feedback perturbation in English (left) and Spanish (right). Asterisks indicate statistical significance at <0.001 (***) and marginal 0.053 (*).

Adaptation magnitude significance (i.e., difference from 0) was assessed with one-tailed two-sided t-tests in R. English-dominant speakers showed significant positive adaptation in English (t = 8.4046, df = 325, p = <0.001) but non-significant adaptation in Spanish (t = 1.6628, df = 328, p = 0.097). Spanish-dominant speakers showed non-significant adaptation in English (t = -0.068297, df = 297, p = 0.9456) and non-significant adaptation in Spanish (t = 1.0525, df = 320, p = 0.29). Linear mixed effects modelling using *lmerTest* [8] in R was used to determine the statistical significance of the differing adaptation magnitudes across the two task languages and speaker groups. *Task language (TL)*, *dominant language (DL)* and *task language*dominant language (TL*DL)* were fixed effects, with *participant* as a random effect. TL of Spanish was associated with significantly lower adaptation magnitude (t = -4.827, df = 1250, p = < 0.001). There was a significant interaction between TL and DL (t = 5.078, df = 1252.325, p = <0.001). Pairwise comparisons were calculated using *emmeans* [11] and corrected using the multivariate t-distribution method. English-dominant speakers showed greater adaptation response magnitude in English than in Spanish (t = 4.218, df = 1308, p = 0.0001). Spanish-dominant speakers showed greater adaptation response magnitude in English than in Spanish, although the difference was marginally non-significant (t = 2.42, df = 1254.4, p = 0.053).

3.2. Categorical vowel perception task results

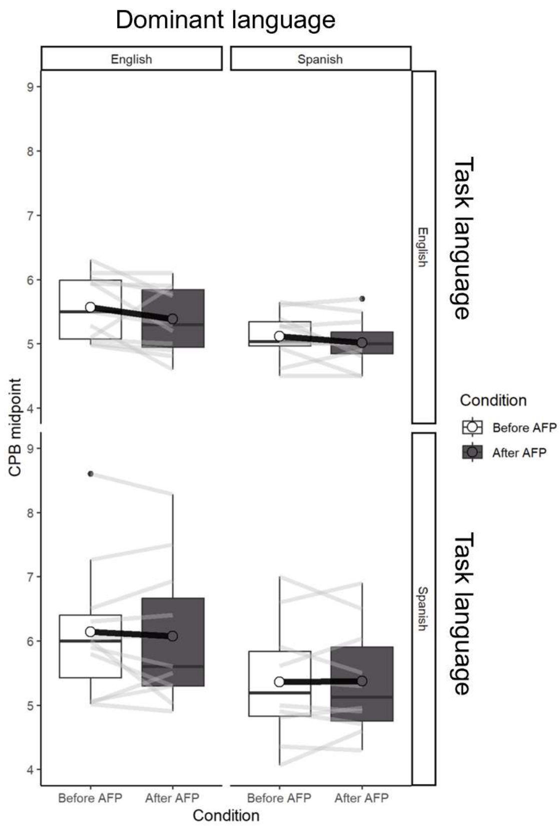

Figure 3: Left: Categorical perceptual boundary midpoints in English before and after feedback perturbation in Spanish-dominant group, in English (above) and Spanish (below). Right: Categorical perceptual boundary midpoints in Spanish before and after feedback perturbation in English-dominant group, in English (above) and Spanish (below).

Following calculation of category perceptual boundaries (CPB) for each continuum, participant and test phase, linear mixed effects modelling was used to assess the statistical significance of the change i n the perceptual boundary location before and after AFP in each language (see Fig. 3). Fixed effects were *condition* (before or after AFP), with participant as a random effect. Overall, the effect of condition was non-significant ($t = 0.536$, $df = 63.21264$, $p = 0.594$). Ultimately, even one robust adaptation response was not associated with statistically significant change in the location of the categorical boundary.

4. DISCUSSION

Only one feedback perturbation condition, English-dominant speakers producing English, led to expected changes in formant frequencies indicative of adaptation. This result suggests that while compensatory responses to altered feedback are apparent for vowels in the dominant language, the direction of adaptation is less reliable in a second language, and less reliable for an otherwise comparable vowel in Spanish than in English.

These differences between Spanish and English support previous research in suggesting that adaptation is stronger at category boundaries [12]. Unlike English-dominant speakers, Spanish speakers showed no significant adaptation in their dominant language; although formant changes were in the expected direction, effect sizes were significantly reduced and non-significant. This result hints that the vowel inventory of the task language may influence adaptation, alongside language dominance. The English vowel inventory is denser than the Spanish vowel inventory: in a more crowded vowel space, the same feedback perturbation may have a more pronounced effect since greater sensitivity to auditory targets and boundaries must be used to maintain differences between vowel categories. An effect of density also supports a result in which both English and Spanish speakers opposed the perturbation in their dominant languages, but only the English speakers did so significantly.

In none of our conditions did we observe any reliable change in categorical perception boundaries, even though participants produced robust adaptation responses in one of the conditions (English-dominant speakers producing English words). These null effects suggest that perceptual change previously shown to stem from speech motor learning may be restricted to the perturbed vowel itself and not applied to a similar vowel in the surrounding vowel space.

This result has implications for the perception-production relationship. Post-AFP after-effects in speech production ([4],[9],[10]) indicate that adaptation to perturbation persists beyond the removal of altered feedback. Meanwhile, the lack of after-effects in speech perception observed here, in contrast to monolingual results in [9], suggest either: (1) that perception is not necessarily altered by a change in production, and therefore that speech production representations are not a key source of information in speech perception, or (2) that the underlying representations are sufficiently category-specific that transfer between even highly similar vowels in L1 and L2 is absent. Further research will be required to disentangle these two possibilities.

It is possible that the experimental characteristics and task order (English tasks first) had a detrimental effect on adaptation in the later Spanish-language tasks, and that this lack of adaptation response was associated with a subsequent lack of change in the categorical boundary. Adaptation after-effects from feedback perturbation can linger throughout the washout phase [13] and may have required a more targeted task to ensure that the after-effects of the perturbation were fully eradicated before starting the new block.

In order to correct for this possibility, we are currently collecting more data with a different task design: categorical perception data is only collected in one language (the unperturbed language) with one week in between different language sessions. This weeklong break between AFP sessions should ensure that adaptation after-effects are completely washed out before starting a new task. We are also collecting more data with an improved counterbalancing, where the first perturbed language (English or Spanish) varies for odd and even-numbered participants. This improved experimental design will allow us to more accurately capture and assess cross-language transfer in perception and production and determine whether these processes are present.

5. CONCLUSIONS

The findings shown here provide evidence for a mediating role of task language in adaptation to formant perturbations. Furthermore, this experiment did not find a significant effect on categorical perception boundaries after a feedback perturbation task in the other language. This result held regardless of task language or the speaker's dominant language. It is possible that this result arises due to a disconnection between production change and perceptual boundary change when the test crossed languages, even in a condition where adaptation magnitude was significantly higher than 0. These findings suggest that changes in production of one vowel do not automatically entail changes in categorical perception of nearby vowels in another language.

6. REFERENCES

[1] Birdsong D., Gertken L. M., Amengual M. (2012, January 20). Bilingual Language Profile: An easy-to-use instrument to assess bilingualism. *COERLL, University of Texas at Austin [online]*. URL: https://sites.la.utexas.edu/bilingual/

[2] Brainard, D. H. 1997. The Psychophysics Toolbox. Spatial Vision 10, 433-436.

[3] Cai, S., Boucek, M., Ghosh, S.S., Guenther, F.H., Perkell, J.S. 2008. A system for online dynamic perturbation of formant frequencies and results from perturbation of the Mandarin triphthong /iau/. In: *Proceedings of the 8th Intl. Seminar on Speech Production, Strasbourg, France, Dec. 8 - 12, 2008*, 65-68.

[4] Franken, M. K., Acheson, D. J., McQueen, J. M., Hagoort, P., & Eisner, F. 2019. Consistency influences altered auditory feedback processing. *Quarterly Journal of Experimental Psychology* 72(10), 2371-2379.

[5] Houde, J. F., Jordan, M. I. 1998. Sensorimotor adaptation in speech production. *Science* 279(5354), 1213-1216.

[6] Houde, J. F., & Nagarajan, S. S. 2011. Speech production as state feedback control. *Frontiers in Human Neuroscience*, *5*, 82.

[7] Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., Banno, H. 2008. Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. In: *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, 3933-3936.

[8] Kuznetsova A, Brockhoff P.B., Christensen, R.H.B. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82(13), 1–26.

[9] Lametti, D. R., Rochet-Capellan, A., Neufeld, E., Shiller, D. M., Ostry, D. J. 2014. Plasticity in the human speech motor system drives changes in speech perception. *Journal of Neuroscience* 34(31), 10339-10346.

[10] Martin, C. D., Niziolek, C. A., Duñabeitia, J. A., Perez, A., Hernandez, D., Carreiras, M., Houde, J. F. 2018. Online adaptation to altered auditory feedback is predicted by auditory acuity and not by domain-general executive control resources. *Frontiers in Human Neuroscience* 12, 91.

[11] Lenth, R. 2022. _emmeans: Estimated Marginal Means, aka Least-Squares Means_. R package version 1.8.0, <https://CRAN.R-project.org/package=emmeans>.

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

[13] Parrell, B., Agnew, Z., Nagarajan, S., Houde, J., Ivry, R. B. 2017. Impaired feedforward control and enhanced feedback control of speech in patients with cerebellar degeneration. *Journal of Neuroscience* 37(38), 9249-9258.

[14] R Core Team. 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

[15] Stuart, A., Kalinowski, J., Rastatter, M. P., Lynch, K. 2002. Effect of delayed auditory feedback on normal speakers at two speech rates. *The Journal of the Acoustical Society of America* 111(5), 2237-2241.

[16] Tourville, J. A., & Guenther, F. H. 2011. The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, *26*(7), 952-981.

[17] Tremblay, S., Houle, G., Ostry, D. J. 2008. Specificity of speech motor learning. *Journal of Neuroscience* 28(10), 2426-2434.