SOUND CATEGORIZATION AFTER SPEAKING WITH A BITE BLOCK

Xinyu Zhang, Rob Schoonen, Esther Janse

Radboud University
xinyu.zhang@ru.nl, rob.schoonen@ru.nl; esther.janse@ru.nl

ABSTRACT

Earlier research has investigated how speech sound categorization is affected by displacement of the articulators during listening. We investigated perceptual behavior *following* a manipulation to restrict tongue raising, and tested whether the experience of having spoken with a bite block affects subsequent categorization of an /ɪ/-to-/ɛ/ continuum. Furthermore, we investigated whether being able to hear one's own manipulated production affects this potential perceptual recalibration (by including a condition with noise-masking during production). Surprisingly, participants in the bite block group gave fewer /ɛ/ responses than the no bite block group at pretest, but response patterns were similar for the two groups at posttest. This suggests that anticipating speaking with a bite block affected categorization behavior more than the actual speech experience. As such, the results do not provide evidence that sound representations change due to articulator displacement.

Keywords: speech perception, perceptual recalibration, speech production

1. INTRODUCTION

The much-debated motor theory of speech perception [1] proposed that part of perceiving speech sounds is perceiving the articulatory movement (but cf. e.g., [2, 3] for critical notes). Later empirical studies, using e.g., 'virtual-lesion' methods like rTMS [4]), showed that disruption of premotor cortex during perception impairs speech perception. But can changes in speech production experience induce changes in speech perception?

Several studies have perturbed the articulators *while* participants perform a perception task in order to investigate the interaction between speech production and perception. Yeung and Scott [5] had participants breathe either through the mouth or nose while giving nasality judgments to auditory stimuli. Participants were more likely to judge a sound as 'nasal' when breathing through the nose. Similarly, Trudeau-Fissette et al. [6] applied facial stretch in the direction of the mouth shape as if producing the vowel /e/, and observed that participants perceived more stimuli as /e/ when there was skin stretch, as compared to no skin stretch.

Another study investigated the subsequent perceptual effect *following* a manipulation of the articulators [7]. Sato and colleagues investigated perceptual categorization of a noise-masked /pa/ vs. /ta/ before and after articulator training, with half of their participants getting lip training (repeatedly protruding lips as if making a kiss-like gesture) and the other half tongue training (repeatedly pressing tongue to palate). Following training, the 'lip group' was biased towards perceiving /p/, whereas the 'tongue group' was biased towards /t/. Whereas Sato and colleagues engaged speakers' articulators, their training did not involve speech production.

Other studies manipulated speakers' auditory feedback during a speech task [8, 9]. Using altered auditory feedback (AAF), Shiller and colleagues [8] upshifted speakers' auditory feedback of centre of gravity of /s/, inducing a compensatory downshifted /s/. Subsequent sound categorization aligned with the altered auditory input during speaking: those who had had the AAF gave more /s/ responses than at pretest, which was not observed for a control group (but cf. [9] for different results).

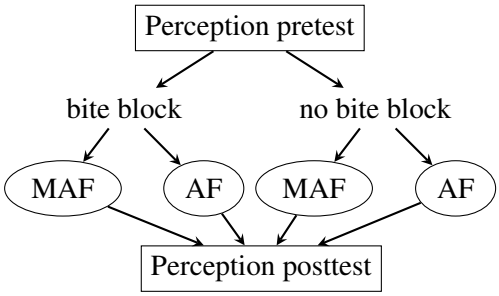
The current study used a bite block to alter speech production, impairing tongue height distinctions. More specifically, we tested for a potential recalibration in perceptual categorization of a vowel height contrast. Importantly, as our focus was on *subsequent* rather than *simultaneous* effects of articulator displacement, the bite block was removed during perceptual testing. Additionally, in line with earlier research on what drives perceptual change [8, 9], we isolated the effect of hearing one's altered speech from the motor experience associated with the altered production, by noise-masking the auditory feedback during speech production for half of our participants. Thus, we address the research question of whether having full access to auditory feedback of one's own altered speech matters. Lastly, if the bite block mainly affects response bias rather than perceptual representations per se [7], the bite-block effect might be strongest for the most ambiguous continuum steps.

2. METHODS

2.1. Experimental Design

Sound categorization was compared between pretest and posttest and was always tested without bite block obstruction. A production task was administered between pretest and posttest, in which participants read aloud a set of pseudowords from a computer screen (see §2.3.1). Participants were randomly assigned to a bite-block or no-bite-block (control) group. Within each of the two groups, participants would either have access to normal auditory feedback of their own productions, or would hear speech-shaped noise to mask their production. Figure 1 illustrates the experimental design[1].

Figure 1: Experimental design. Ellipses represent the production task; rectangles represent perception tasks; MAF: masked auditory feedback; AF: regular auditory feedback.



2.2. Participants

Sixty participants aged between 18 and 30 years were recruited through the Radboud Research Participation System. They were randomly assigned to one of the four experimental Biteblock-Auditory Feedback conditions (see Figure 1). All were native speakers of Dutch, with normal or corrected-to-normal vision, and they all reported having normal hearing and speech abilities (i.e., none reported stuttering or any other speech problems), and none reported dyslexia. They were all compensated for their participation.

2.3. Materials

2.3.1. Production stimuli

The target vowels for the perception and production tasks were /ɪ/ and /ɛ/. The production stimuli were bisyllabic pseudowords conforming to Dutch phonotactics. The syllable structure for the pseudowords (targets and fillers alike) was C1VC2-/ɪf/, such that the target vowel is always embedded in a stressed CVC structure (e.g., /ˈtɪsɪf/, /ˈtɛsɪf/). The addition of the standard second syllable made the pseudowords less 'wordlike'. Care was also taken that no C1VC2 combination (either in targets or in fillers) formed real Dutch morphemes to further minimize lexical frequency effects. For target stimuli, in order to allow easy segmentation of the target vowels (for the production analysis not reported here), C1 and C2 were voiceless plosives and fricatives. Filler stimuli followed the same syllable structure, and contained more diverse vowels in the stressed V position, as well as more diverse consonants in the C1 and C2 positions with other manners of articulation. The places of articulation of C1 were evenly distributed in the target items: /p/, /t/, and /k/ each appeared twice. Each of the stop consonants was paired with each of the vowels (3∗2 = 6 target stimuli). In addition to the six target stimuli, the production set contained 12 filler stimuli, all repeated three times [(6 targets + 12 fillers)∗3 repetitions = 54 items for production].

2.3.2. Masking noise

Speech-shaped noise was used as masking noise for the masking-noise-as-auditory-feedback condition. Speech-shaped noise was created by filtering random noise with a filter made by averaging the long-term average spectrum of 10 male and 10 female adult speakers reading a phonetically balanced text. The resulting 25-minutes-long noise file, with a sampling frequency of 44100 Hz, was played during the entire production block for those in the masked auditory feedback (MAF) groups.

2.3.3. Perception Stimuli

The perception pretest and posttest blocks contained three continua of seven steps of the target contrast /ɪ/ to /ɛ/, as well as nine filler continua of seven steps [(3 targetContrasts + 9 fillerContrasts) ∗ 7 stepsEach = 84 items in total]. The end points of each (target or filler) continuum were a monosyllabic real-word minimal pair in Dutch (e.g., [pɪp]-[pɛp]). Each step of each continuum was presented once in each perception block. Three types of filler continua were included: different vowel formant continua (3 pairs, e.g., [biːr] - [byːr]), consonant place of articulation continua (3 pairs, e.g., [tau] - [kau]), and voice onset time continua (3 pairs, e.g., [toːs] - [doːs]).

2.4. Procedures

The experiment was written and run in Open Sesame [10]. All procedures were carried out at the Centre for Language Studies Lab of Radboud University in one experimental session that took less than 20 minutes in total. Upon arrival in the lab, participants were informed that they were assigned to the bite-block group or to the control (no bite-block) group. Before the start of the perception pretest, participants who were assigned to the bite block group received instructions on how to insert the bite block. Instructions included verbal description (keep tongue flat under the bite block, where to place the incisors), and being shown a photo of the experimenter with the bite block in the mouth. Importantly, participants did not put the bite block in their mouths, nor tried speaking with it, before the perception pretest.

The bite block used in the experiment was an unopened plastic bottle of a probiotic drink (75 mm tall, smallest circumference 24 mm, largest circumference 38 mm), with the original 65 ml of product still inside, see Figure 2. Each participant assigned to the bite-block group used a new bottle which they could take with them afterwards.

In each perception block, participants were presented with tokens from different contrast continua and asked to decide which of the words of the real-word minimal pair they heard. At the start of each trial, two options (the two words of the pair) appeared on the screen 500 ms before the audio was presented. After the initial 500ms, the two options remained on the screen while an auditory token (one step from the corresponding auditory continuum) was played through a pair of Sennheiser HME 110 headphones. Participants were then asked to indicate which word they heard, by pressing either the left or right key on a button box, corresponding to the left or right option on the screen, respectively[2]. The order of the stimuli was randomized per participant and per block.

The production task was self-paced by key press. The pseudoword prompts were displayed on a BenQ PD2700U computer screen (font size 32, resolution 1920 * 1080). Audio recordings were made using a Sennheiser ME 64 microphone and stored per token. Participants went through a practice phase of five items for the perception task before completing the perception pretest. The order of the production stimuli was randomized per participant and per block. Masking noise during the production block was played through a pair of Avantree ANC 031 active noise canceling headphones.

Figure 2: Bite block used in the present study.

2.5. Statistical Analysis

For each trial, the randomized stimuli and the response by the participant were automatically recorded by the experimental interface (§2.4). All analyses were carried out in R (version: 4.0.4) [11]. A generalized logistic regression mixed-effects model was run on the data using the lme4 package [12], with the perceptual response as the dependent variable. Binary contrasts were coded using treatment coding. Predictor variables include continuum step (7 steps, centered on step 4), auditory feedback condition (levels AF, MAF; with AF mapped on the intercept), bite block condition (levels BB, NB; with NB mapped on the intercept), as well as time of measurement (levels pretest, posttest; with pretest mapped on the intercept). We included random intercepts for participant and minimal pair. The analysis started with the four-way interaction model including all predictor variables[3]: $responseE \sim cstep * prepost * AF * BB + (1|subject) + (1|pair)$. Insignificant effects were taken out in a step-wise manner (taking out higher-order interactions before lower-order ones, followed by removal of simple effects) to arrive at the most parsimonious model. Model comparisons were applied using the anova() function in R after each removal step to verify that exclusion of an effect did not lead to significantly worse model fit.

3. RESULTS

Figure 3 shows categorization behavior at pretest and posttest for bite block and no bite block groups. The most parsimonious model of the likelihood of getting an /ɛ/ response was: $responseE \sim cstep * AF * BB + prepost * BB + AF * BB + prepost * AF + (1|subject) + (1|pair)$. Model estimates are provided in Table 1.

As expected, participants were more likely to give an /ɛ/ response for the higher steps on the continuum (β = 1.970 (0.158), $p < 0.001$), and the no bite block group (on the intercept of the logistic regression model) did not change their categorization behavior from pretest to posttest (β = -0.368 (0.258), *n.s.*). At pretest, participants in the bite block group were less likely to give an /ɛ/ response than the no

Figure 3: Percentage of /ɛ/ response at pretest and posttest for bite block group (BB) and no bite block group (NB) at each continuum step (collapsing over auditory feedback conditions).

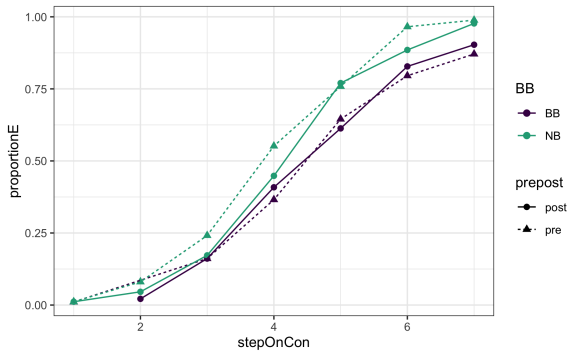
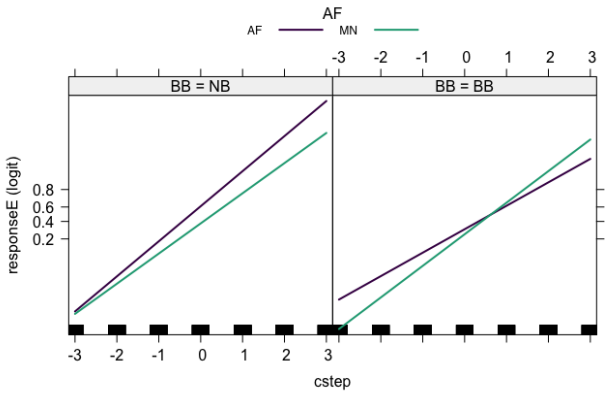


Table 1: Coefficients for sound categorization model; β coefficients represent logits.

	β	SE	z value	p value
(Intercept)	0.632	0.608	1.035	0.300
cstep	1.969	0.158	11.834	<0.001
MAF	-0.820	0.604	-1.357	0.175
BB	-1.603	0.595	-2.675	0.007
posttest	-0.368	0.258	-1.420	0.153
cstep:BB	-0.653	0.179	-3.637	<0.001
cstep:MAF	-0.276	0.196	-1.407	0.159
MAF:BB	0.682	0.818	0.833	0.405
BB:posttest	0.625	0.280	2.229	0.026
cstep:MAF:BB	0.733	0.255	2.870	0.004

bite block group (β = -1.603 (0.595), $p < 0.05$). The significant interaction between posttest and bite block condition ($\beta = 0.625$ (0.280), $p < 0.05$) shows that the difference between the groups is smaller at posttest. This was verified by running the same model with posttest (rather than pretest) mapped on the intercept, showing no bite block effect at posttest. For the no bite block group, masking noise did not affect categorization behavior at the center of the continuum (β = -0.820 (0.604), *n.s.*). This null effect of masking noise also holds for the other continuum steps (β = -0.276 (0.196), *n.s.*). For the bite block group, there was an effect of masking noise on categorization behaviour, but only at the more extreme continuum steps (as evident from the triple interaction in Table 1, and as visualized in Figure 4, $\beta = 0.6733$ (0.255), $p < 0.01$). However, as these effects held at pretest and posttest alike, they cannot be attributed to (noise-masking effects on) the actual bite-block production experience, but rather seem to reflect pre-existing group differences between the two bite-block groups (cf. design overview in Figure 1).

Figure 4: Model effects plot showing the three-way interaction (MN: masking noise; AF: regular auditory feedback).



4. DISCUSSION

We set up a sound categorization study with a pretest-posttest design to investigate whether tongue height restriction due to speaking with a bite block affects subsequent categorization of a vowel height contrast continuum. Our second question was whether being able to hear one's own altered speech would affect this potential shift. Noise masking during production was used to address this second point.

Regarding the first research question, the bite block and no bite block groups were expected to differ at posttest, if at all, rather than at pretest. Had our results aligned with studies that tested perception during articulation manipulation [5, 6], those in the bite block group would give more /ɛ/ responses at posttest, as the bite block fixes the tongue to a more /ɛ/-like configuration. Recall that instruction about the bite block occurred before pretest, without speakers actually trying it themselves. It is unclear whether these pretest results would replicate. If so, we can only speculate that those anticipating speaking with tongue restriction mapped their /ɪ/ onto a more open position, hence the 'regular' /ɛ/ might seem more /ɪ/-like to them. However, the observation of no bite block group difference at posttest does not provide reliable evidence that sound representations change due to articulator displacement.

Now that our results have not shown a reliable perceptual shift driven by articulator displacement, our second research question does not apply anymore. Follow-up research is planned to investigate bite block and noise masking effects on speech acoustics, and to test whether individual amount of acoustic change in production predicts perceptual change.

5. REFERENCES

[1] A. M. Liberman and I. G. Mattingly, "The motor theory of speech perception revised," *Cognition*, vol. 21, no. 1, pp. 1–36, 1985.

[2] G. Hickok, "The role of mirror neurons in speech and language processing," *Brain and Language*, vol. 112, no. 1, p. 1, 2010.

[3] A. J. Lotto, G. S. Hickok, and L. L. Holt, "Reflections on mirror neurons and speech perception," *Trends in Cognitive Sciences*, vol. 13, no. 3, pp. 110–114, 2009.

[4] I. G. Meister, S. M. Wilson, C. Deblieck, A. D. Wu, and M. Iacoboni, "The essential role of premotor cortex in speech perception," *Current Biology*, vol. 17, no. 19, pp. 1692–1696, 2007.

[5] H. H. Yeung and M. Scott, "Postural control of the vocal tract affects auditory speech perception." *Journal of Experimental Psychology: General*, vol. 150, no. 5, p. 983, 2021.

[6] P. Trudeau-Fisette, T. Ito, and L. Ménard, "Auditory and somatosensory interaction in speech perception in children and adults," *Frontiers in Human Neuroscience*, vol. 13, p. 344, 2019.

[7] M. Sato, K. Grabski, A. M. Glenberg, A. Brisebois, A. Basirat, L. Menard, and L. Cattaneo, "Articulatory bias in speech categorization: Evidence from use-induced motor plasticity," *cortex*, vol. 47, no. 8, pp. 1001–1003, 2011.

[8] D. M. Shiller, M. Sato, V. L. Gracco, and S. R. Baum, "Perceptual recalibration of speech sounds following speech motor learning," *The Journal of the Acoustical Society of America*, vol. 125, no. 2, pp. 1103–1113, 2009.

[9] D. R. Lametti, A. Rochet-Capellan, E. Neufeld, D. M. Shiller, and D. J. Ostry, "Plasticity in the human speech motor system drives changes in speech perception," *Journal of Neuroscience*, vol. 34, no. 31, pp. 10 339–10 346, 2014.

[10] S. Mathôt, D. Schreij, and J. Theeuwes, "Opensesame: An open-source, graphical experiment builder for the social sciences," *Behavior Research Methods*, vol. 44, no. 2, pp. 314–324, 2012.

[11] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2022. [Online]. Available: https://www.R-project.org/

[12] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[1] The flow chart presented here was part of a larger experimental design, which included more production blocks and a production analysis.

[2] The option containing '[ɪ]' did not always appear on the left. The '[ɪ]' option was shown on one side of the screen for two of the minimal pairs and on the other side for the other minimal pair.

[3] responseE: the likelihood of getting an /ɛ/ response; cstep: continuum step, centered on step 4 (continua always go from /ɪ/ to /ɛ/ from step 1 to step 7); AF: auditory feedback condition; BB: bite block condition; pair: minimal pair continuum.