THE ROLE OF SOMATOSENSATION IN PERCEPTUAL RECALIBRATION FROM SPEECH IMAGERY

Jacob B. Phillips*, Lenore A. Grenoble*, Giovanna Hooton*, Peggy Mason**

*Department of Linguistics; **Department of Neurobiology, University of Chicago
correspondence address: jbphillips@uchicago.edu

ABSTRACT

Perceptual recalibration is the lasting effect of adjustments in category boundaries as a result of non-auditory stimuli, including speech imagery [12], i.e. silent and imagined articulations. The present study examines perceptual recalibration from speech imagery in American English sibilants and asks what role haptic feedback and the recruitment of the motor system play in perceptual recalibration. We compare a control group to Kim, a woman with a unique congenital neuropathy who lacks all somatosensation. The results of this study find evidence for perceptual recalibration from both silent and imagined articulations in the control group, but no evidence for recalibration in either condition for Kim. These findings suggest that perceptual recalibration from speech imagery may require the activation of haptic sensations even when the articulators are immobile.

Keywords: perceptual recalibration, speech imagery, haptic feedback, sensory neuropathy

1. INTRODUCTION

Speech perception is a dynamic process, with category boundaries between speech sounds constantly shifting, often in response to non-auditory stimuli. Perceptual recalibration is the aftereffect of those adjustments, resulting from exposure to an ambiguous sound paired with non-auditory information that biases categorization in one direction or another.

Lexical factors bias listeners to hear words over non-words – the Ganong effect [6] – which persists after exposure to induce perceptual recalibration. Norris et al. [9] demonstrated that following exposure to repetitions of the Ganong effect, listeners are more likely to categorize an ambiguous sound as belonging to the category that yielded a lexical word in the exposure period. Similar patterns of recalibration have been observed for phonotactic information, with listeners biased toward sounds that resulted in licit clusters [5]; visual information, with listeners biased toward sounds reinforced by visual articulations [3]; and reading, with listeners biased toward sounds represented orthographically [7].

Perceptual recalibration has also been induced by the actions of the listener. Shiller [13] demonstrated that following exposure to altered auditory feedback, participants not only exhibit shifts in production, but also in perception; yet participants exposed to shifted auditory stimuli without producing simultaneous articulations exhibit no such recalibration. These findings suggest a role of haptic feedback and/or motor planning in inducing recalibration. Similarly, Scott [12] demonstrated that Arabic listeners exhibit recalibration following exposure to ambiguous /b/-/d/ stimuli when they silently articulate one end of the continuum in sync to the auditory stimuli, suggesting a role of feedback/planning even when the listener is overtly aware that their articulations are not producing the auditory signal. Furthermore, Scott observed a similar effect when the speech is merely imagined, with listeners exhibiting recalibration following saying one of the sounds in their head without moving any articulators.

For perceptual recalibration from speech articulation or speech imagery, it is not immediately clear whether distributional learning resulting from the activation of a phonetic category alone is enough to induce recalibration [8], or whether the motor system is recruited through speech planning, haptic feedback or activated haptic sensations. Forward models of speech perception suggest that an individual's motor system is only consulted as needed, like when dealing with ambiguous sounds [10] or speech imagery [14]. This results in the activation of haptic sensations in the brain, even for imagined speech where no haptic feedback is received.

The present study examines Kim, an individual with a unique congenital neuropathy that means she receives no haptic feedback from articulation. We ask if she exhibits any differences in patterns of perceptual recalibration as a result of mouthed or imagined speech compared to a control group with no reported neurological impairments. As imagined speech has been demonstrated to yield motor planning and activation of haptic sensations [14], no difference is predicted between the imagined or mouthed speech for either Kim or the control group.

If category activation alone induces recalibration, this predicts Kim should exhibit recalibration in both mouthed and imagined speech. If haptic feedback and haptic sensations are necessary to induce recalibration, this predicts Kim should not exhibit recalibration from either mouthed or imagined speech.

2. CASE STUDY: KIM

Kim is a thirty-nine-year-old American female with a severe one-of-a-kind variant of Hereditary Sensory Autonomic Neuropathy (HSAN) Type II [1]. She lacks both small- and large-fiber somatosensory afferents on her body, head, and oral cavity. She has no tracheal sensation and does not cough in response to liquids entering her trachea. However, Kim has intact motor nerves and muscle strength. For example she can cough deliberately. Kim is motorically limited due to her lack of sensory feedback. She has no reflexes and cannot stand or walk independently nor can she chew, spit or suck.

Kim had maxillary hypoplasia, mandibular prognathism, frontal bossing and a "fish mouth" at birth. At age seventeen, Kim underwent successful facial reconstruction surgery to bring out the maxilla and correct a Class III malocclusion by reducing the mandibular protrusion. A LeFort 1 osteotomy, bilateral sagittal split ramus osteotomy, and horizontal anterior osteotomy of the mandible were performed. Kim attended speech therapy as a child, yet she remains unable to produce a complete bilabial constriction and her bilabials are noticeably non-standard at slower speech rates.

Without somatosensation, Kim is presented with a unique challenge for speech production: she receives no tactile feedback from a constriction or bracing of the tongue body, no proprioceptive information about the position of her articulators, and no sensation of pressure building behind a constriction. This suggests that for Kim speech planning must involve exclusively acoustic rather than articulatory goals and bars the activation of haptic sensations during speech imagery and perception.

3. METHODS

The present study was designed to test the role of haptic feedback in perceptual recalibration from speech imagery, building off the findings of Scott [12]. For this experiment, there were two speech imagery conditions: 'enacted' (MOUTHED) and 'pure' (IMAGINED). There were two word conditions (*sin* and *shin*), yielding a 2 X 2 design. All participants were tested on all conditions in alternating blocks.

3.1. Participants

Twenty-nine undergraduate students (mean age 20) at the University of Chicago were recruited for the control group and received course credit for their participation. All participants were native speakers of American English. Thirteen participants identified as female, fifteen as male, and two as genderqueer. No participants reported neurological or speech disorders/abnormalities. One additional participant took part in the study but was not included in the analysis due to non-attentive responses.

3.2. Stimuli

A twenty-one step continuum from /s/ to /ʃ/ was created from natural speech produced by a twenty-seven-year-old male speaker of American English. The sibilant onsets from *sin* and *shin* were extracted and digitally mixed using a custom Praat [4] script with twenty-one different scaling ratios, including the extreme endpoints (0%[s]–100%[ʃ] and 100%[s]–0%[ʃ]). The manipulated sibilant onsets were then cross-spliced onto a separate token of *sin*, creating a continuum from *sin* to *shin*.

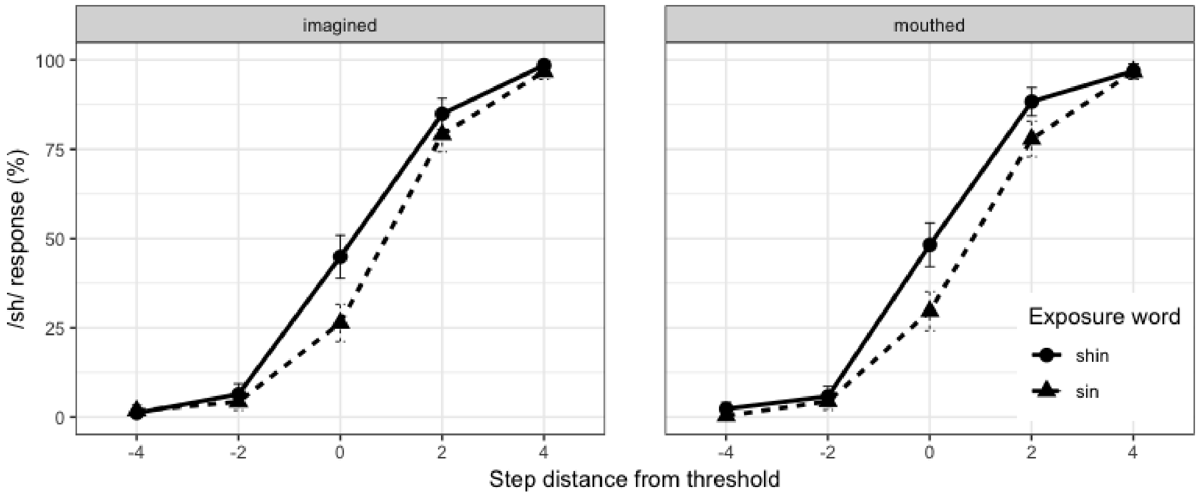
3.3. Procedure

Experimentation for the control group took place in an isolated double-walled sound booth, while Kim participated in a quiet wheelchair accessible room. All participants were seated in front of a monitor and fitted with Sennheiser HD 555 headphones.

In a pretest, an individual threshold between /s/ and /ʃ/ was identified for each participant. Each step of the continuum was played 4 times binaurally and participants responded by typing either 0 or 1 corresponding to an orthographic representation on the monitor. Key assignment was counterbalanced between subjects. A potential threshold was identified as a step that received two categorizations as both /s/ and /ʃ/. If multiple steps exhibited split categorization, the median was selected as the threshold.

After a threshold was selected, participants alternated between four repetitions of each of the four conditions. Each block was subdivided into exposure and test phases. In the exposure phase, participants were exposed to eight repetitions of their identified threshold between /s/ and /ʃ/ with 1500 ms silence between repetitions. A pulsing yellow-to-red circle was presented in the middle of the screen, timed to the repetition of the audio tokens. In the MOUTHED condition, participants were instructed to 'silently mouth [*sin* or *shin*] to beat of the pulsing circle' and were told to think of it as 'lip syncing'.

Figure 1: Mean and 95% confidence interval for percent *shin* responses on a five step continuum centered on each individual's threshold (x-axis) as a function of the speech imagery (panels) and repeated word (line/shape).



In the IMAGINED condition, participants were asked to 'say [*sin* or *shin*] in [their] head without moving their mouth, lips or tongue.' To avoid confusion between the different tasks, before the experiment began participants practiced the different tasks with the researcher present using a [b]*ear* to [p]*ear* continuum. In order to ensure that participants were accurately implementing the intended conditions for each block, participants responded to each prompt by typing the intended word (`sin` or `shin`) and image (`mouth` or `imagine`).

Immediately following the exposure phase, there was a 2500 ms pause after which participants were asked to categorize two repetitions of a five-step continuum, defined as their threshold plus two and four steps above and below their threshold. The setup of the test phase was identical to the pretest.

3.4. Analysis

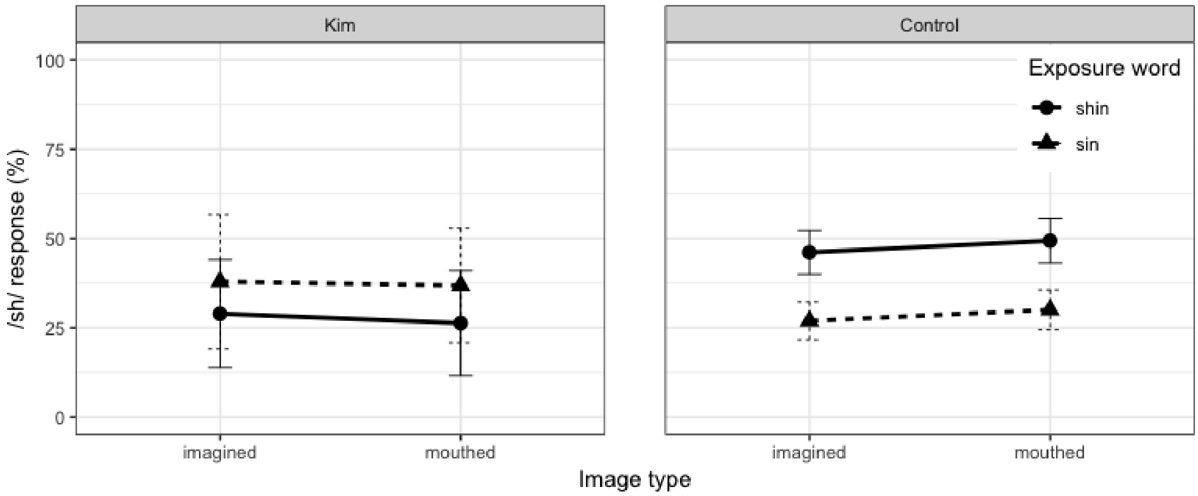
Kim's responses (/s/ or /ʃ/) were modeled using a generalized linear model with the `glm()` function in R [11]. The model included trial ORDER (scaled), WORD (SIN, SHIN), IMAGE (IMAGINED, MOUTHED), STEP (-4–4, scaled) and the immediately preceding PRIORSTEP (-4–4, scaled) as independent variables. The control group responses were modeled with the same explanatory variables using a generalized linear mixed model with the `glmer()` function [2]. Additionally, random intercepts for subject and ORDER and by-subject random slopes for WORD and IMAGE improved model likelihood, suggesting significant individual variability.

4. RESULTS

There were initial concerns that in the absence of visual feedback, e.g. from a mirror, Kim would have difficulty consistently producing silent articulations without feedback of any kind: the task necessitated the lack of auditory feedback and Kim has no access to haptic feedback. Nevertheless, Kim produced silent articulations in the MOUTHED condition consistently to the beat of the pulsing circle. However, in the IMAGINED condition, Kim occasionally exhibited unintended labial and possibly lingual movements, which appeared to be significantly reduced compared to the MOUTHED condition. None of the subjects in the control group were observed to produce silent articulations in the IMAGINED condition.

Beginning with the control group, Figure 1 plots the proportion of *shin* responses, illustrating a noticeable difference in stimuli categorization depending whether individuals repeated *sin* or *shin* in both the IMAGINED and MOUTHED conditions. Specifically, regardless of image condition, individuals were more likely to categorize their threshold token as *shin* if the word repeated was *shin* than if it was *sin*. These observations were supported by the model with a significant increase in *shin* responses if the repeated WORD was *shin* ($z = 3.64, p < 0.001$) in addition to a significant increase in *shin* responses as higher (i.e. more-/ʃ/) steps ($z = 35.31, p < 0.001$). This appears to extend to a lesser degree in tokens two steps above individuals' thresholds in the MOUTHED condition but not the IMAGINED condition, however, this was not supported by the model

Figure 2: Mean and 95% confidence interval for percent *shin* responses at individuals' thresholds as a function of speech imagery (x-axis) and repeated word (line/shape) for Kim (left panel) and the control group (right panel).



with no difference between image conditions.

Turning to Kim's responses, Figure 2 compares Kim's responses at her identified threshold to those of the control group. Visual inspection does not suggest any difference in stimulus categorization as a result of the repeated word in either the IMAGINED or MOUTHED condition. Likewise, the model for Kim's responses predicts no significant role of WORD or IMAGE condition. Unlike the control group, Kim appears to exhibit a trial effect, with an increased likelihood of *shin* responses as the experiment progresses.

5. CONCLUSIONS

The present study demonstrates that neither enacted (MOUTHED) nor pure (IMAGINED) speech imagery induces perceptual recalibration for an individual who congenitally lacks all haptic feedback. In contrast, both forms of speech imagery induce recalibration for a control group who reported no neurological impairments, replicating and extending the findings of Scott [12] to American English sibilants. This contrast suggests that haptic feedback and activated haptic sensations, rather than speech sound category activation or motor planning, may play a crucial role in recalibration from speech imagery.

If Kim were to have exhibited recalibration, this would have demonstrated that haptic sensations are not required to induce perceptual recalibration. With no observed recalibration, however, it is unclear whether recalibration requires somatosensation or whether Kim does not exhibit recalibration for reasons independent of her sensory neuropathy, as significant individual variation was observed in the control group, including a small minority of individuals who, like Kim, show no effect of the repeated word. If her neuropathy is the root of her observed pattern, it is additionally unclear at present whether this is due to her contemporaneous lack of feedback during articulation or a developmental effect, such that she never received haptic feedback during language acquisition in order to later activate echos of those sensations during speech imagery and speech perception.

In addition, the recalibration effect appears to be skewed toward /s/: responses following exposure to *sin* are significantly biased away from /ʃ/ responses, but responses following exposure to *shin* remain at 50%. This asymmetric distribution may be attributed to the stimuli creation, as the onset sibilant was cross-spliced onto a naturally produced token of *sin*, which may contain coarticulatory information biasing an /s/ response. This suggests that perceptual recalibration is stronger when the repeated word aligns with the inherent coarticulatory biases of the stimuli [9]. Otherwise, this asymmetry may suggest that listeners are more familiar with variation in /s/ and thus more susceptible to recalibration from silent articulations of /s/ than /ʃ/.

6. ACKNOWLEDGMENTS

We would like to thank Kim and her family for their collaboration. This work is supported by the Neubauer Collegium at the University of Chicago.

7. REFERENCES

[1] Axelrod, F. B., Gold-von Simson, G. 2007. Hereditary sensory and autonomic neuropathies: types II, III, and IV. *Orphanet Journal of Rare Diseases* 2(39).

[2] Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Stastical Software* 67(1), 1–48.

[3] Bertelson, P., Vroomen, J., de Gelder, B. 2003. Visual recalibration of auditory speech identification: A McGurk aftereffect. *Psychological Science* 14(6), 592–597.

[4] Boersma, P., Weenink, D. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5(9/10), 341–345.

[5] Cutler, A., McQueen, J. M., Butterfield, S., Norris, D. 2008. Prelexically-driven perceptual retuning of phoneme boundaries. *Proceedings of Interspeech* 2056.

[6] Ganong, W. F. 1980. Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception & Performance* 6(1), 110–125.

[7] Keetels, M., Schakel, L., Bonte, M., Vroomen, J. 2016. Phonetic recalibration of speech by text. *Attention, Perception & Psychophysics* 78(3), 938–945.

[8] Kleinschmidt, D. F., Jaeger, T. F. 2016. Re-examining selective adaptation: Fatiguing feature detectors, or distributional learning? *Psychonomic bulletin & review* 23(3), 678–691.

[9] Norris, D., McQueen, J. M., Cutler, A. 2003. Perceptual learning in speech. *Cognitive Psychology* 47(2), 204–238.

[10] Pickering, M. J., Garrod, S. 2007. Do people use language production to make predictions during comprehension? *Trends in cognitive sciences* 11(3), 105–110.

[11] R Core Team, 2015. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna.

[12] Scott, M. 2016. Speech imagery recalibrates speech-perception boundaries. *Attention, Perception & Psychophysics* 78(5), 1496–1511.

[13] Shiller, D. 2009. Perceptual recalibration of speech sounds following speech motor learning. *Journal of the Acoustical Society of America* 125(2), 1103–1113.

[14] Tian, X., Poeppel, D. 2010. Mental imagery of speech and movement implicates the dynamics of internal forward models. *Frontiers in Psychology* 1, 1–23.