INDIVIDUALS' PERCEPTUAL RETUNING PREDICTS ARTICULATORY ACCOMMODATION: IMPLICATIONS FOR SOUND CHANGE

Patrice Speeter Beddor[1], Andries Coetzee[1], Ian Calloway[2], Stephen Tobin[3] and Ruaridh Purse[1]

[1]University of Michigan, MI, USA, [2]ALEX-Alternative Experts, VA, USA, [3]Universität Potsdam, Germany
beddor@umich.edu, coetzee@umich.edu, icalloway@alexinc.com, stephen.tobin@uni-potsdam.de, rupurse@umich.edu

ABSTRACT

Individuals differ from each other, as speakers and listeners, in the extent to which they adapt to unfamiliar speech patterns. The hypothesis that listener-specific perceptual adjustments for an unfamiliar pattern are reflected in the listener-turned-speaker's imitation of the pattern was tested for raising of English /æ/ before /g/ (e.g., [beɪg] *bag* but [bæk] *back*). For 37 American English participants, perceptual learning and spontaneous imitation of raised /æ(g)/ were assessed using eye-tracking and ultrasound imaging, respectively. Results support the hypothesis that perceptual retuning predicts, in part, articulatory accommodation: the more an individual perceptually adapted to raised /æ(g)/ (e.g., used [beɪg] to rapidly disambiguate *back-bag* trials), the more that individual imitated raised /æ(g)/. These findings are viewed as relevant to the spread of (coarticulatorily motivated) change in that community members who attend particularly closely to innovative interlocutors' novel forms may be especially likely to converge towards those forms in their subsequent productions.

Keywords: sound change, imitation, perceptual learning, production-perception link

1. INTRODUCTION

Speakers' productions and listeners' percepts are malleable and dynamic. For example, speakers modify their speech in ways that converge towards the patterns of an interlocutor [18] or a model talker [10, 22]. Listeners also adapt; their perceptual decisions show sensitivity to community speech norms [12] and to the patterns of a specific idiolect [16]. Yet despite malleability, adaptation is highly variable across individuals: speakers differ in how much and how accurately they imitate [14] and listeners differ in the extent to which they learn a new phonetic contrast or native-like perceptual weights [9, 20]. These individual differences have been found to be reliable, with individual differences in degree of convergence towards a talker, for example, being stable across test sessions [25].

This study tests the hypothesis that an individual's perceptual adjustments for a novel, coarticulatory speech pattern will be reflected in their imitation of that pattern, that is, that perceptual retuning predicts, in part, articulatory accommodation. Theoretically, our approach is motivated by our interest in testing accommodation in the laboratory as a vehicle for understanding the spread of coarticulatorily motivated change in a speech community. In particular, we aim to better understand the phonetic behaviors of early adopters of an incipient change in a speech community—community members who, arguably, are especially likely to converge towards an innovative interlocutor's novel forms. Empirically, the expectation of a relation between an individual's production and perception of coarticulated speech is consistent with recent evidence that individuals who produce more extensive coarticulation (e.g., vowel nasalization) are also more efficient users of that information [3, 4, 27, 28]. Yet not all studies find a comparable production-perception relation [11, 20]. Moreover, when perception and production involve adaptation to an unfamiliar pattern, there are social, attitudinal, and other factors [1, 2] that might mitigate against the hypothesized link.

The targeted novel pattern is raising of /æ/ towards [eɪ] before /g/ (but not /k/; as in [beɪg] *bag* vs. [bæk] *back*), a vowel shift found for speakers in parts of the Northern U.S. and Canadian Prairies [21]. Mielke et al.'s [17] ultrasound study of English speakers who do and do not produce raised /æ/ before /g/ (henceforth, raised /æ(g)/) showed that, even for non-raisers, the tongue root was more advanced and tongue body more fronted for /g/ than /k/ following /æ/, consistent with enlarging the supralaryngeal cavity to sustain voicing during /g/. This more fronted dorsal constriction for /g/ presumably underlies, or has at least contributed to, /æ(g)/ raising.

In this study, participants unfamiliar with this pattern were exposed, using a visual world task, to a model talker's raised /æ(g)/ (and unraised /æ(k)/) (see [5, 23] for a similar approach). Participants' spontaneous imitation of that pattern was assessed in a subsequent ultrasound session. This approach allowed us to study the time course of participants' reorganization of their perceptual and articulatory spaces as they learn an unfamiliar phonetic variant. Our main interest is in the *relation* between perceptual and articulatory learning and reorganization for individual listener-speakers.

2. PERCEPTUAL ADAPTATION

The perception task assesses whether, over the course of an eye-tracking experiment, learning novel raised /æ(g)/ facilitates participants' recognition of (unraised) /æk/-final words but slows recognition of (phonetically similar) /eɪk/-final words.

2.1. Methods

Auditory stimuli for the audio-visual trials were 10 sets of minimal C(C)VC triplets where V = /æ/ or /eɪ/ and final C = /g/ or /k/ (e.g., *back-bag-bake*) produced by a model talker with native /æ(g)/ raising. This talker's diphthongal /æ(g)/ had an F1/F2 trajectory similar to that of /eɪ(k)/ but with a higher F1 frequency. Because /æ/ was longer before /g/ than /k/, minor stimulus editing was done to reduce the average /æ(g)/-/æ(k)/ difference to about 35 ms (and thus to reduce duration as information for the voicing contrast). Visual stimuli were black and white line drawings corresponding to each word. Each audio-visual stimulus for a trial consisted of one auditory stimulus and two side-by-side visual images, one target (corresponding to the auditory stimulus) and one competitor. Visual targets and competitors were all six possible combinations of stimuli for a triplet (e.g., *back-bag*, *back-bake*, *bake-bag*, *bake-back*, *bag-back*, *bag-bake*). Auditory stimuli were presented in a blocked design in which words for the first 30 trials ended only in /æk/ or /eɪk/. Following this pre-exposure block was a 150-trial exposure block containing all target words, including /æg/-final words. Across the full experiment, each audio-visual pairing occurred three times.

Participants were 37 listeners whose variety of American English does not exhibit /æ(g)/ raising. For each trial, participants heard the recorded instruction "Look at the pictures;" 2 s later a fixation cross appeared on the computer screen and participants heard "Fixate cross. (pause) Now look at [target word])." Eye movements over the course of each trial were monitored using an EyeLink 1000 Plus (SR Research) remote eye-tracker.

2.2. Predictions

Trials with target /Cæg/ were learning trials; our main predictions are for trials with competitor /Cæg/, which test effects of that exposure. For target *back* | competitor *bag*-type trials, participants should fixate on *back* earlier during the exposure block than during pre-exposure because /æ(g)/ raising should reduce *back-bag* competition. The reverse pattern (later target fixation) should hold for target *bake* | competitor *bag*-type trials because of increased *bake-bag* competition due to raised /æ(g)/. A further prediction is that individual listeners will differ in whether, and the extent to which, the novel pattern influences their lexical decisions.

2.3. Results

Analyses of participants' fixations were conducted using Generalized Additive Mixed Modeling in R [19], as implemented in the *mgcv* [26] and *itsadug* [24] packages. Target fixation proportions from 200-1200 ms after the target vowel onset were extracted for analysis, allowing 200 ms to plan and launch a saccade [15]. The fixation proportions (binary responses) were computed with a logit link function with the following model structure: experimental block (pre-exposure / exposure), stimulus competitor type, and elapsed time over the course of a trial (as well as their interactions) were modeled as fixed effects; by-participant random smooths for elapsed time were also included.

The modeled fixation results in Figure 1, left panel, show that, when the target was /Cæk/, that is, when raised /æ(g)/ provided early disambiguating information during the vowel portion of the trial (*back-bag*), listeners fixated earlier and more often on the target during the exposure block than during pre-exposure. However, when the target was /Ceɪk/ (Figure 1, right), that is, when /æ(g)/ raising increased ambiguity (*bake-bag*), fixations were later and less frequent during the exposure block. Thus, aggregate results support predictions (see also [23]).

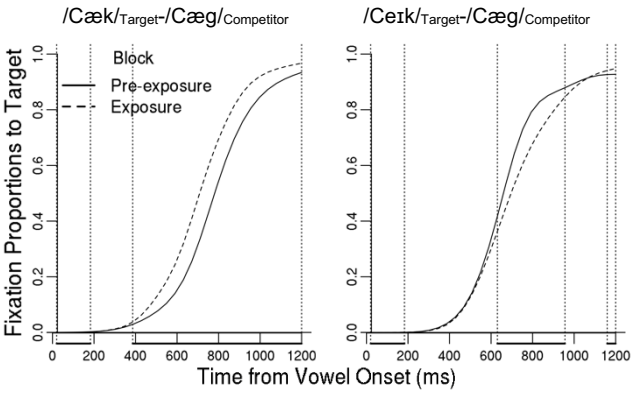

Figure 1: Perception: model-derived fixations, according to block, on /Cæk/ (left) and /Ceɪk/ (right) targets when competitor = /Cæg/. (Dotted lines: region of significant difference between blocks)

To quantify *individual* differences in perceptual retuning, we calculated, for each listener, the average difference between exposure and pre-exposure blocks in time to first target fixation for trials with /Cæg/ as the visual competitor. In this paper, we focus on target /Cæk/ | competitor /Cæg/ trials where, the larger the difference score, the greater the reduction in lexical competition for that listener—and, arguably, the

greater the learning. The median difference score was 43 ms, with 30% of participants' scores falling within ±20 ms of the median. However, scores ranged from -46 ms to 203 ms, indicating that some participants' looks to target were substantially faster after hearing raised /æ(g)/ whereas those of others slowed somewhat. In subsequent models (section 3), each participant's "difference to first target fixation" score serves as the *perception score* used to test our hypothesis that perceptual adjustments for a novel speech pattern are predictive of individuals' imitation of that pattern.

3. ARTICULATORY ACCOMMODATION

3.1. Methods

The same 37 participants returned about 1-2 weeks later for the production study. Auditory and visual stimuli were those used in the perception study. Production trials were of two types and used an alternating talker-participant naming task (see [6]). In "talker" trials, a talker icon appeared with the stimulus image, indicating that the participant should simply listen to the image name (e.g., *Say bag*) produced by the model talker. In "participant" trials, a microphone icon appeared with the stimulus image, indicating that the participant should name the image.

Table 1 gives the structure of the experiment. In blocks 1 (pre-exposure) and 3 (post-exposure), the participant produced words of all three types but did not hear the talker say /Cæg/ words. In block 2 (exposure), both participant and talker produced all word types. To maximize imitation, in all blocks half of the trials were repetition trials in that the word the participant was prompted to produce was the same as the immediately preceding word spoken by the talker.

	/Cæg/		/Cæk/		/Ceɪk/	
Block	T	P	T	P	T	P
1 (pre-exposure)	-	40	40	20	40	20
2 (exposure)	60	100	50	30	50	30
3 (post-exposure)	-	40	40	20	40	20

Table 1: Number of each trial type produced by the model talker (T) and participant (P) in the imitation study, according to block.

Tongue contour data were collected using a Zonare z.one ultrasound system with a p4-1c phased-array transducer. The probe was held in place by a custom-made, lightweight ultrasound stabilizer [7].

3.2. Results

Ultrasound sequences were submitted to MTracker [29] for automated tongue contour extraction. Vowel raising (i.e., imitation) is assessed by comparing, across blocks, the frame-wise maximum y-coordinate (normalized within participants) of participants' tongue contours from vowel onset to offset. These y-coordinate values were submitted to a GAMM in R again using the *mgcv* [26] and *itsadug* [24] packages. To investigate our hypothesis that perceptual and articulatory accommodation are linked at the level of the individual, the model included as fixed effects the Perception Scores (difference in first fixation for /Cæk/-/Cæg/ eye-tracking trials; section 2.3), the interaction factor of Rhyme (/æg/, /æk/, /eɪk/) x Block, and smooths for Normalized Time x Perception Score for each level of Rhyme x Block. Random smooths over Normalized Time for each level of Participant x Rhyme x Triplet were also included. Only /æg/ results are reported here.

Figure 2, left panel, gives the modeled aggregate results for maximum normalized tongue height for /æ(g)/ across the time course of the vowel for the three blocks. As would be expected for pre-velar /æ/, the tongue dorsum is highest at vowel offset (at onset of /g/ closure). However, across much of the duration of the vowel, tongue body height raises with increasing experience with the model talker's raised /æ(g)/ (from blocks 1 to 2 and 3).

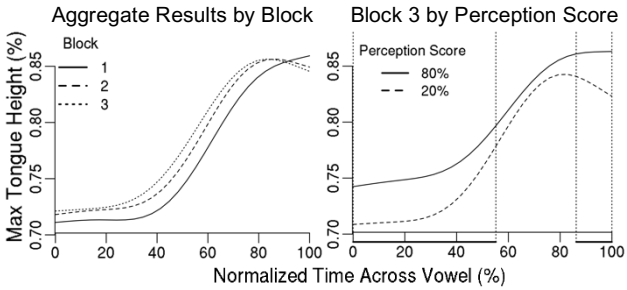

Figure 2: Production: model-predicted maximum tongue height. Left: aggregate results according to block. Right: block 3 (post-exposure) results for participants who adapted more (80th percentile) and less (20th) in perception.

Figure 2, right, assesses whether individuals' perceptual retuning predicts imitation and gives, for the post-exposure block, modeled /æ(g)/ tongue height values for individuals who fall near the 80th and 20th percentiles of the perception scores. Consistent with a retuning-imitation link, maximum tongue height was higher across the initial half of the vowel for individuals who adapted more in perception (80th percentile) than those who adapted less (20th).

Although Figure 2 shows that stronger perceptual adapters had overall higher /æ(g)/ realizations at the end of the task, further inspection of imitation results by block reveals a more nuanced pattern. Figure 3 shows that perceivers who adapted more (top panels) imitated quickly (from blocks 1 to 2, left) and then

plateaued (blocks 2 to 3, right) whereas those who adapted less (bottom) showed increasing tongue height across the task.

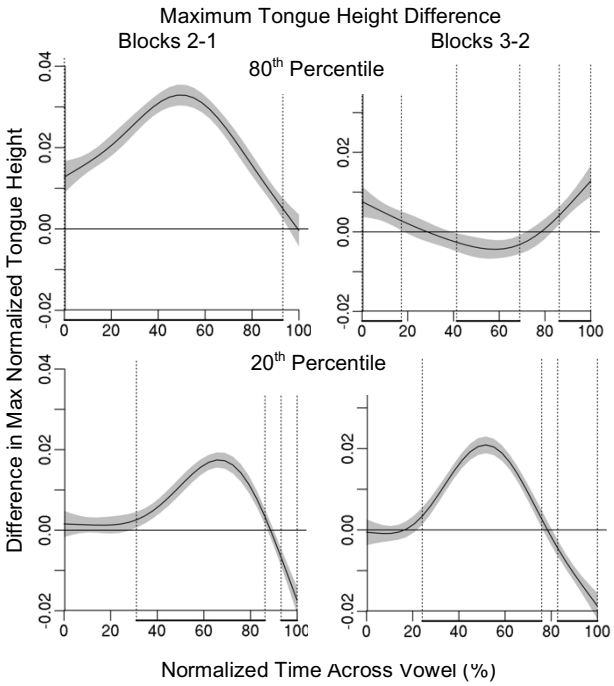

Figure 3: Model-predicted differences in maximum tongue height between Blocks 2-1 (left panels) and Blocks 3-2 (right panels) for participants whose perception scores fell near the 80th (top) and 20th (bottom) percentiles.

4. DISCUSSION

This study investigated American English-speaking participants' perceptual learning of, and articulatory accommodation to, an unfamiliar, coarticulatorily motivated pattern of raised /æ/ before voiced velar /g/. Consistent with previous work [5, 23], listeners exposed to this pattern over the course of an eye-tracking task showed, on average, faster recognition of minimal pair competitors ending in /æk/ (e.g., *back* vs. *bag*) but slower recognition of minimal pair competitors ending in /eɪk/ (*bake* vs. *bag*), indicating that, as listeners learned more information about the model talker's vowel patterns, they perceptually recalibrated in ways that influenced their lexical decisions in real time. These same participants also accommodated to this pattern as speakers, with ultrasound imaging data showing that average maximum tongue body height for /æ(g)/ was higher after participants had been exposed, in a "take turn" spontaneous imitation task, to the same model talker's raised /æ(g)/.

Despite these clear aggregate patterns, individual participants differed from each other, both as listeners and speakers, in degree of accommodation to the model talker. (Responses to a post-test survey provide no evidence that these individual differences in accommodation can be attributed to differences in participants' prior familiarity with the novel pattern.) We hypothesized that the perceptual and articulatory differences are linked: that an individual's degree of perceptual retuning for a novel pattern would predict, to some extent, their imitation of that pattern. In support of this hypothesis, the results of a production model in which each participant's degree of perceptual learning (for target /Cæk/ | competitor /Cæg/ trials) was included as a predictor for imitation showed that, the more an individual perceptually adapted to raised /æ(g)/ (i.e., used raised /æ(g)/ to rapidly disambiguate *back*-*bag*-type trials), the more that individual imitated raised /æ(g)/. Indeed, these stronger perceptual adapters were both more accurate imitators (more closely approximating the model talker's yet more raised /æ(g)/) and more rapid imitators, exhibiting especially large imitative tongue height positions during the exposure block.

That the degree of an individual's perceptual accommodation to a novel pattern is linked to their articulatory accommodation may seem unsurprising from the perspective that articulatory accommodation has perceptual and articulatory components: the novel pattern must be a sufficiently salient perceptible difference to be imitated. These data, though, indicate that the link is tighter than "sufficiently salient" and rather show that perceptual *retuning* is predictive of imitation.

We conclude by considering the implications of these findings for theories of the spread of sound change. To illustrate, we take the incipient change of spread of raised /æ(g)/ through a speech community. Mielke et al.'s [17] data show that even speakers from communities without this pattern tend to produce fronted /g/ relative to /k/ constrictions. Thus, English-speaking communities will tend to have non-random directional /g/ ~ /k/ variation (with /g/ having a more front articulation) and, at least in some cases, co-occurring directional /æ/ variation (higher tongue body before /g/ than before /k/). A challenge for sound change researchers is to explain why the more innovative forms of this systematic variation (i.e., especially fronted /g/ and raised /æ(g)/) might become the new phonetic norm in a speech community (e.g., [8], [13]). By hypothesis, our data shed some light on the phonetic behaviors of early adopters of these more innovative raised /æ(g)/ variants—and, more generally, of early adopters of innovative coarticulatorily motivated phonetic forms. The findings raise the possibility that these early adopting listeners-turned-speakers are perhaps especially attentive to the new variants, adjust their perceptual space (or repertoire) accordingly, and mirror those adjustments in their own, now converging productions.

5. ACKNOWLEDGMENTS

This material is based on work supported by NSF Grant BCS-1348150 to Patrice Beddor and Andries Coetzee; any opinions, findings, and conclusions are the authors' and do not necessarily reflect the views of the NSF.

6. REFERENCES

[1] Babel, M. 2010. Dialect divergence and convergence in New Zealand English. *Language in Society* 39, 437-456.

[2] Babel, M. 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics* 40, 177-189.

[3] Beddor, P.S., Coetzee, A. W., Boland, J. E., McGowan, K. B., Styler, W. 2018. The time course of individuals' perception of coarticulatory information is linked to their production: implications for sound change. *Language* 94, 1-38.

[4] Coetzee, A. W., Beddor, P. S., Styler, W., Tobin, S., Bekker, I., Wissing, D. 2022. Producing and perceiving socially structured coarticulation: coarticulatory nasalization in Afrikaans. *Laboratory Phonology* 13, 1-43.

[5] Dahan, D., Drucker, S. J., Scarborough, R. A. 2008. Talker adaptation in speech perception: adjusting the signal or the representations? *Cognition* 108, 710-718.

[6] Delvaux, V., Soquet, A. 2007. The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica* 64, 145-173.

[7] Derrick, D., Carignan, C., Chen, W. R., Shujau, M., Best, C. T. 2018. Three-dimensional printable ultrasound transducer stabilization system. *J. Acoust. Soc. Am.* 144, EL392-EL398.

[8] Garrett, A., Johnson, K. 2013. Phonetic bias in sound change. In: Yu, A. C. L. (ed), *Origins of Sound Change: Approaches to Phonologization*. Oxford University Press, 51-97.

[9] Golestani, N., Zatorre, R. J. 2004. Learning new sounds of speech: reallocation of neural substrates. *Neuroimage* 21, 494-506.

[10] Goldinger, S. D. 1998. Echo of echoes? An episodic theory of lexical access. *Psychological Review* 105, 251-279.

[11] Grosvald, M. 2009. Interspeaker variation in the extent and perception of long-distance vowel-to-vowel coarticulation. *Journal of Phonetics* 37, 173-188.

[12] Hay, J., Warren, P. Drager, K. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34, 458-484.

[13] Harrington, J., Kleber, F., Reubold, U., Schiel, F., Stevens, M. 2018. Linking cognitive and social aspects of sound change using agent-based modeling. *Topics in Cognitive Science* 10, 707-728.

[14] Honorof, D. N., Weihing, J., Fowler, C. A. 2011. Articulatory events are imitated under rapid shadowing. *Journal of Phonetics* 39, 18-38.

[15] Hutton, S. B. 2008. Cognitive control of saccadic eye movements. *Brain and Cognition* 68, 327-340.

[16] Kraljic, T., Brennan, S. E., Samuel, A. G. 2008. Accommodating variation: dialects, idiolects, and speech processing. *Cognition* 107, 54-81.

[17] Mielke, J., Carignan, C., Thomas, E. R. 2017. The articulatory dynamics of pre-velar and pre-nasal /æ/-raising in English: an ultrasound study. *J. Acoust. Soc. Am.* 142, 332-349.

[18] Pardo, J. S. 2006. On phonetic convergence during conversational interaction. *J. Acoust. Soc. Am.* 119, 2382-2393.

[19] R Core Team. 2021. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. http:/www.R-project.org/

[20] Schertz, J, Cho, T., Lotto, A., Warner, N. 2015. Individual differences in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics* 52, 183-204.

[21] Stanley, J. A. 2022. Regional patterns in prevelar raising. *American Speech* 97, 374-411.

[22] Tobin, S. J. 2022. Effects of native language and habituation in phonetic accommodation. *Journal of Phonetics* 93, 101148.

[23] Trude, A. M., Brown-Schmidt, S. 2012. Talker-specific perceptual adaptation during online speech perception. *Language and Cognitive Processes* 27, 979-1001.

[24] van Rij, J., Wieling, M., Baayen, R. H., van Rijn, H. 2020. itsadug: interpreting time series and autocorrelated data using GAMMs. R package version 2.4.

[25] Wade, L., Lai, W., Tamminga, M. 2021. The reliability of individual differences in VOT imitation. *Language and Speech* 64, 576-593.

[26] Wood, S. 2019. mgcv: mixed GAM computation vehicle with automatic smoothness estimation. R-package version 1.8–31.

[27] Yu, A. C. L. 2019. On the nature of the perception-production link: individual variability in English sibilant-vowel coarticulation. *Laboratory Phonology* 10, 1–29

[28] Zellou, G. 2017. Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics* 61, 13–29.

[29] Zhu, J., Styler, W., Calloway, I. 2019. A CNN-based tool for automatic tongue contour tracking in ultrasound images. *arXiv preprint arXiv:1907.10210.*