OCTOBER 28 2025

Expectation-driven shifts in perception and production

Lacey Wade; Meredith Tamminga

J. Acoust. Soc. Am. 158, 3517–3528 (2025)
https://doi.org/10.1121/10.0039577

JASA ARTICLE

Expectation-driven shifts in perception and production

Lacey Wade[1,a)] and Meredith Tamminga[2,b)]
[1]Department of Linguistics, University of Kansas, Lawrence, Kansas 66045, USA
[2]Department of Linguistics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

ABSTRACT:
While phonetic convergence has been taken as evidence for tight perception–production links, attempts to correlate perceptual adjustments with production shifts have been inconsistent, and the existence of expectation-driven convergence further complicates our understanding of this relationship. Here, we report the results of a go/no-go lexical decision task showing that expectation-driven perceptual shifts occur toward the same stimuli that has previously been shown to elicit expectation-driven convergence. We also replicate previous expectation-driven convergence results in production using the Word Naming Game [from Wade (2022). Language **98**(1), 63–97]. However, we fail to find evidence that individuals' expectation-driven shifts in perception correlate with those in production. Findings are discussed in terms of implications for the role of expectations on linguistic behavior and the relationship between perception and production. © *2025 Acoustical Society of America*. https://doi.org/10.1121/10.0039577

(Received 9 January 2025; revised 24 September 2025; accepted 30 September 2025; published online 28 October 2025)
[Editor: Molly Babel] Pages: 3517–3528

I. INTRODUCTION

A. The perceptual underpinnings of phonetic convergence

Phonetic convergence occurs when individuals shift their speech patterns to align with those of another talker, spontaneously, and usually with little awareness. Convergence has often been taken as empirical evidence for a tight connection between linguistic perception and production mechanisms (e.g., Goldinger, 1998; Pickering and Garrod, 2013; Olmstead *et al.*, 2013), with observed shifts in production being attributed to reproduction of the perceptual input. A perceptual shift—the result of perceptual adaptation, where novel pronunciations are integrated into the percept of a given phone such that the intended phone can be accurately retrieved from this mapping—is therefore a reasonable prerequisite to expect for a parallel production shift. After all, how can a novel pronunciation for a given phone be produced if the novel pronunciation was not accurately integrated into the percept for that phone? However, attempts to correlate perception and production shifts have been inconsistent and unreliable (Pardo, 2012; Schertz *et al.*, 2023; Kim and Clayards, 2019), calling into question just how close that connection may be.

The perceptual underpinnings of convergence may be difficult to observe for a number of reasons, including disruption of the perception–production link through mediating processes that inhibit or facilitate automatic shifts. Still, accounts of convergence generally assume *some* automatic perceptual component, in part because convergence has been observed in relatively asocial laboratory situations lacking obvious social motivations to converge (e.g., Shockley *et al.*, 2004; Goldinger, 1998). Much recent work on convergence therefore takes a "hybrid" approach, positing that both automatic mechanisms and social motivations (and potentially other mediating mechanisms) play a role (e.g., Pardo *et al.*, 2017; Babel, 2012; Walker and Campbell-Kibler, 2015; Pardo, 2012). Observations that convergence is both linguistically and socially selective provide evidence that any perception–production link involved in convergence is not entirely direct; if convergence were a fully automatic consequence of perception, it should occur consistently across features and social contexts (Pardo, 2012). In terms of linguistic selectivity, attempts at contrast preservation have been suggested to inhibit convergence (Nielsen, 2011; Kim and Clayards, 2019), explaining why imitation of lengthened-Voice Onset Time (VOT) is more commonly observed than that of shortened-VOT, which would obscure the voiced–voiceless stop contrast in English (although cf., Mitterer and Ernestus, 2008). Other work has found that convergence tends to occur toward features that are different enough from one's own to motivate converging, but also similar enough that their own production repertoire allows it (Walker and Campbell-Kibler, 2015). In terms of social selectivity, attitudes toward the model talker (Yu *et al.*, 2013; Babel, 2010, 2012), social awareness and evaluations of the feature (Clopper and Dossey, 2020; Lee-Kim and Chou, 2024; Walker and Campbell-Kibler, 2015; Clopper *et al.*, 2024), and the semantic content of the utterance (Yu, 2013; Babel, 2010) have all been shown to mediate degree and/or direction of convergence.

Hybrid accounts positing mechanistic perception–production links that are inhibited/enhanced by external factors still predict *some* relationship—at minimum, an implicational one, wherein perception shifts are necessary, but not sufficient, for production shifts (e.g., Kraljic *et al.*, 2008).

[a)]Email: laceywade@ku.edu
[b)]Email: tamminga@ling.upenn.edu

Further evidence for an implicational relationship comes from comparisons between explicit convergence (where participants are explicitly instructed to imitate a model talker) and implicit convergence (where participants are not instructed to converge), which have shown that just because somebody *can* converge when prompted, does not mean they will when not prompted. For instance, Schertz (2025) finds that explicit convergence correlates with discrimination accuracy in perception, while spontaneous convergence does not. This suggests that implicit convergence leaves substantial room for mediating factors, such as the tendency to focus on the relevant acoustic dimension when not prompted or social motivations that may inhibit convergence. Such variability may obscure the link observed for explicit imitation, since explicit instructions may override social motivations to imitate and reduce variability in attention paid to the speech signal. In sum, while spontaneous convergence may deviate from perception for varying reasons, it is still assumed that imitation relies on perception in some way—after all, how can you imitate something you did not perceive?[1]

B. Expectation-driven convergence

Expectation-driven convergence complicates this picture by demonstrating that people can imitate variants that are socially cued and therefore *expected* from a talker, even when those variants are not directly observed within the interaction. Expectation-driven convergence is a robust phenomenon, attested in both naturalistic field recordings (Fasold, 1972; Bell, 2001; Auer and Hinskens, 2005) and in the lab (Wade, 2020; Wade and Roberts, 2020; Wade, 2022; Wade *et al*., 2023). In earlier work, expectation-driven convergence was generally observed in interactional settings and appealed primarily to social motivations. For instance, Fasold (1972) observed that speakers in Washington, DC, produced more African American English (AAE) features when conversing with an African American interviewer compared to a White interviewer. Similarly, Bell (2001) reported that an Anglo interviewer in New Zealand frequently used the *eh* tag—a feature stereotypically associated with male Māori speech—when conversing with a male Māori interviewee but not when conversing with an Anglo interviewee, even though the Māori interviewee never actually used this feature. To account for such observations, Auer and Hinskens (2005) proposed the "Identity-Projection model" of convergence, which "does not mean imitating the actual speech of one's co-participant, but rather conforming to some stereotyped image of how a person in the social role of the co-participant ought to, or can be expected to, behave" (p. 343). Recent work has extended evidence for expectation-driven convergence to laboratory settings (e.g., Wade, 2020, 2022; Wade *et al*., 2023), including in an artificial "alien language" (Wade and Roberts, 2020). Most relevant for our purposes is Wade (2022), which we closely replicate here. After listening to a U. S. Southern-accented talker who never produced any instances of the /aɪ/ vowel, participants shifted their own speech to produce more monophthongal /aɪ/, consistent with expectations for this feature in the Southern dialect. Dialect background influenced both the strength and trigger of convergence: Southern participants converged more than non-Southerners (Wade, 2022), and in a follow-up study, non-Southerners converged toward a talker labeled as "Southern" (even if they were in fact from Ohio), while Southerners only converged when they observed acoustic cues consistent with Southern-shifted speech (Wade *et al*., 2023).

When individuals converge toward features not directly derived from the immediate input, it requires an explanation that does not rely on the direct reproduction of phonetic properties of the input. Indeed, earlier work observing expectation-driven convergence tended to be rooted in socio-psychological accounts of convergence like Communication Accommodation Theory (Giles *et al*., 1991) and Audience Design (Bell, 1984), positing primarily social motivations. If speakers can shift their speech toward a linguistic target they did *not* just hear, what does this mean for the perception–production link in phonetic convergence—and by extension, our evidence base for such a link in general?

There is a possible explanation for expectation-driven convergence that *does* maintain a central role for the perception–production link. In our prior work, we have suggested that experimental participants who exhibit expectation-driven convergence are generating a new production target from their social expectations: when they hear a Southern accent, they update their own production targets to be more congruent with the cluster of dialect features they are hearing. However, an alternative possibility is that the social expectations induced by the accent influence the participants' *perceptual* expectations, and then the updated perceptual categories in turn trigger a shift in the production target. There is ample evidence that perceptual shifts of various kinds can be induced by social information, including expectations (e.g., Strand and Johnson, 1996; Niedzielski, 1999; D'Onofrio, 2015, 2018; Koops *et al*., 2008), supporting the plausibility of such an account. We are not aware of any attempts to establish an empirical connection between expectation-driven perceptual adaptation and expectation-driven convergence.

C. Expectations and perceptual adaptation

It is empirically well-established that linguistic changes in perception can be induced by non-linguistic information, supporting the plausibility of an expectation-driven perceptual adaptation account. For example, Strand and Johnson (1996) found that participants who were visually cued to believe they were listening to a female compared to a male talker perceived different boundaries between /s/ and /ʃ/, consistent with expectations that women have higher spectral frequencies for fricatives than men. Niedzielski (1999) found that, when speakers believed they were listening to a Canadian speaker, they chose raised-diphthong tokens as

https://doi.org/10.1121/10.0039577

representative of the /aʊ/ diphthong, but when they thought they were listening to a Detroit speaker they did not, since the raised /aʊ/ diphthong is stereotypically associated with Canadian but not with Detroit speakers. Social expectations can affect lexical access as well. D'Onofrio (2015) found that participants primed to hear "Valley Girl" speech, which is associated with /æ/-retraction, were more likely to look at and click on /æ/ words like *sack* compared to /ɑ/ words like *sock* when exposed to stimuli with vowels ambiguous between /æ/ and /ɑ/. We take these types of perceptual adjustments to be akin to those observed after exposure to a locally observed novel pronunciation, usually from the same talker (e.g., an /s/ sound ambiguous between /s/ and /ʃ/, causing the boundary between these two sounds to shift toward /ʃ/) (see Samuel and Kraljic, 2009 for an overview), despite being triggered by different types of cues.

Notably, different types of adjustment patterns have been reported in the literature on adaptation to novel dialect features. On the one hand, adjustments may be directional such that shifts in word endorsement or continuum classification occur only in the direction of the exposed change. On the other hand, adjustments may involve general laxing of category boundaries, such that novel pronunciations are generally more accepted, regardless of directionality. Zheng and Samuel (2020) suggest that directional recalibration of phonemic boundaries does not seem to play a role in accent accommodation, but rather that accent accommodation is accomplished by relaxing phonemic categorization criteria. However, Bissell and Clopper (2025) find that whether adaptation involves directional shifts or general category broadening varies based on experience level, with less experienced listeners favoring broadening and more experienced listeners shifting their boundaries, but only when the direction is consistent with their experience. Babel *et al.* (2021) report similar findings, suggesting that exposure to /z/-devoicing yields directional adaptations because it is a familiar pronunciation, while /s/-voicing yields general category broadening since this shift is quite unexpected. Ultimately, they argue that both strategies are possible but depend on various properties of the stimulus.

Maye *et al.* (2008) found that inducing learning by passively exposing participants to a passage in which all of the front vowels had been lowered resulted in increased identification of novel words (not heard in the passage) with synthetically lowered front vowels as "words" in a lexical decision task. This method of measuring perceptual learning has also been used to investigate perceptual responses to expected forms. Using a similar paradigm, Weatherholtz (2015) found that exposing participants to a novel chain shift resulted in greater "word" responses for items with shifted vowels in a lexical decision task. Importantly, listeners exposed to only a subset of the chain shift were able to fill in the gaps and generalize to phonemes that were not present in the training phase, suggesting listeners had learned a pattern of co-variation among vowel categories. However, the covariation in that case is structural, drawing on listeners' expectations about phonological relationships. We do not know whether listeners would make the same kinds of adjustment for features that might be expected to co-occur because they happen to coexist within the same real-world regional dialect.

It is also unclear whether these kinds of perceptual adjustments make their way into speech production, or conversely, whether convergence stems from such perceptual shifts. If spontaneous convergence occurs toward stimuli that individuals passively observe, we would expect exposure to novel pronunciations in a perceptual learning task to induce production shifts as well, especially since the phenomenon of perceptual learning establishes that a listener is not just passively exposed to a novel pronunciation but that they have somehow integrated it into perceptual categorization processes. Lehet and Holt (2017), for example, found that participants exposed to an artificial English dialect with noncanonical use of f0 for the stop voicing contrast not only decreased their reliance of f0 in perception, but also decreased their own use of f0 in subsequent production, providing some evidence that perceptual learning can influence production. However, Kraljic *et al.* (2008) found that, even though individuals exhibited large perceptual learning shifts in the /s/-/ʃ/ category boundary after exposure to a novel pronunciation, these individuals did not spontaneously exhibit equivalent shifts in production. They concluded that "While such perceptual changes might prove to be necessary for production changes, they do not seem to be sufficient" (p. 15).

Whether *expectation-driven* perceptual shifts make their way into production is even less well understood; we suggest that including expectation-driven phenomena in the question of perception–production relationships may serve to elucidate these inconsistent findings.

D. The present study

The goal of this paper is to compare expectation-driven perceptual adaptation and convergence shifts toward the same stimuli as a test of the "socially-induced percept updating" account of expectation-driven convergence. The same convergence task as we use here has been used successfully in our prior studies (Wade, 2020, 2022; Wade *et al.*, 2023) to elicit expectation-driven shifts in production. We first ask whether sociolinguistic expectations can induce perceptual shifts directly parallel to those we see in an expectation-driven convergence paradigm: does hearing Southern-accented speech make listeners expect to hear more monophthongal /aɪ/, even though they did not hear any form of /aɪ/ in the input? To assess this, we ask whether the same stimuli that induced expectation-driven convergence in Wade (2022) also make participants more likely to accept forms like [braːb] (Southern *bribe*, cf. **brob,* **brab*) as words in a lexical decision task, which would suggest they have made temporary perceptual adjustments to encompass monophthongized /aɪ/. Second, we ask whether such expectation-driven perceptual adaptation is related to changes in production, by replicating the convergence task

from Wade (2022) and correlating individual participants' production and perception results. There are two empirical patterns that we would consider compatible with the idea of expectation-driven convergence being derived from a shift in perception. The strong version of the account, on which the social expectations trigger adjustment of a perceptual category boundary that is shared with the production system, predicts a direct correlation between individuals' perception and production shifts in /aɪ/. A weaker version, on which an adjustment to the perceptual boundary makes the production shift *possible* but not *necessary*, predicts an implicational relationship between individuals' perception and production shifts, such that production shifts should not be able to occur without perceptual shifts allowing for integration of monophthongal /aɪ/ into the /aɪ/ percept.

We do find evidence for perceptual adaptation to encompass monophthongal /aɪ/ tokens, but no evidence that it correlates with—or is a prerequisite for—production shifts toward monophthongal /aɪ/ in production. We discuss possible reasons for the lack of observed correlation, including task effects, multiple co-existing mappings for /aɪ/, and different mechanisms recruited for perception and production.

II. METHODS

A. Participants

We recruited 190 native English speakers through Prolific (Prolific Academic Ltd., London, UK). Participants were excluded based on their performance on the perception and production task separately to maximize sample size within each task, described below.[2] For the combined perception–production analysis, we use data only for the participants that were not excluded from either task, yielding 140 participants for the combined analysis (68 in the Southern condition; 72 in the control Midland condition). Participants were recruited from the U. S., both within the South and outside of the South, with particular recruitment efforts given to Southern participants due to prior evidence that Southerners exhibit greater convergence toward monophthongal /aɪ/.

Perception: Data from 181 participants was analyzed for the go/no-go lexical decision perception task (88 in the Southern condition, 93 in the Midland condition) after excluding those who did not appear to understand the task or complete the task in good faith (i.e., did not press any buttons or pressed a button after every item) ($N = 6$), and participants who did not report their residential history ($N = 3$). We included participants regardless of low accuracy rates for filler and/or non-word items, under the assumption that this reflects the difficulty of the task of identifying Southern-shifted vowels that differ from the non–Southern-shifted productions of most of the participants. For instance, McQueen (1996) suggests that 50% accuracy may be expected on a typical go/no-go task. The Southern accent of the talker in this study may lead us to expect even lower accuracy rates in the present study, so we did not impose any exclusion criteria based on accuracy, other than omitting participants who responded the same way to all fillers and non-words.

Production: Data from 147 participants were analyzed for the production task (73 in the Southern condition, 74 in the Midland condition). Of the 190 participants, 173 had recordings available. Participants were excluded for having unusable data (i.e., due to poor recording quality) for half or more of the elicited words ($N = 23$), or for not providing residential history ($N = 3$). This leaves us with 147 participants.

B. Procedure

The study was coded and administered through Penn Controller for Ibex, supported by MindCore (University of Pennsylvania, Philadelphia, PA) (Zehr and Schwarz, 2018), and participants completed the study through their web browsers. After providing informed consent, participants completed a short demographic survey, providing their age, gender, race, ethnicity, level of education, and residential history. Residential history was used to group participants into "Southern" and "non-Southern" categories, determined based on whether participants spent the majority of their school-aged years (ages 5–18) inside or outside of the *Atlas of North American English* isogloss for /aɪ/-monophthongization (Labov *et al.*, 2006). Participants in the full dataset include 33 non-Southerners/60 Southerners in the Midland condition and 27 non-Southerners/61 Southerners in the Southern condition. Participants then moved on to the experiment, which consisted of a word-naming game production task eliciting expectation-driven convergence, consisting of three phases, with a lexical decision task assessing expectation-driven perceptual shifts embedded between the second and third production phases.

1. Production task

The production task utilized a word-naming game paradigm (Wade, 2020, 2022), in which participants heard or read clues describing various words, then guessed each word aloud using the carrier phrase, "The word is X." For instance, participants might be given the clue, "This is a small, silver U.S. coin worth ten cents," and would respond by stating aloud, "The word is *dime*." When participants were ready to record their response, they pressed a red "Record" button on the screen, and a blinking red dot indicated that they were recording. When they were finished recording, they pressed a "Next" button to continue immediately to the next clue, and the recording indicator turned off. To facilitate accurate responses and reduce data loss, participants were provided on-screen with the number of letters of each correct response, with several letters filled in (e.g., d _ _ e). Each clue was one to two sentences long and, crucially, contained no instances of the /aɪ/ vowel.

The task consisted of three phases: baseline, exposure, and post-exposure. The baseline phase served to collect participants' baseline productions of the /aɪ/ vowel before any exposure to a talker voice. As such, clues were presented on screen, and participants read the clues silently to themselves

https://doi.org/10.1121/10.0039577

before responding aloud. In the exposure phase, participants heard clues read aloud by either a Southern or Midland model talker, depending on the condition to which they had been randomly assigned. The Midland talker was from Youngstown, Ohio and produced typical Midland dialect features; the Southern talker was from Hurley, Mississippi and had recognizable features of the Southern Vowel Shift, including raised front lax vowel nuclei, fronting of back vowels, and the PIN-PEN merger. The model talkers' speech patterns are described in greater detail in Wade (2022). At the start of the exposure phase, the model talker provided auditory instructions (none of which contained the /aɪ/ vowel), which was meant to familiarize participants with the talker's voice. After the exposure phase, participants completed the perception task, described below. Then, they completed the post-exposure phase, where they returned to reading clues on the screen. The post-exposure phase was included to assess how long any convergence effects last post-exposure and to help tease apart whether shifts observed from the baseline to exposure phase were due to the experimental manipulation or due to general fatigue as the experiment progressed.[3]

A total of 180 tokens was elicited from each participant, consisting of 30 target /aɪ/ words and 30 fillers in each phase. Elicited target words all contained the /aɪ/ vowel in coda position (e.g., "fly," "try") or before a voiced consonant (e.g., "dime," "ride"), as this is where /aɪ/ monophthongization most reliably occurs throughout the U.S. South. Monophthongization before voiceless segments (e.g., "light," "rice") is less commonly observed, so these contexts are excluded here. A full list of stimuli can be found in Wade (2022). The order in which words were elicited was randomized across participants, and the phase in which a given word was elicited was counterbalanced across participants. All three sets of 30 target words elicited in a single phase were matched for mean lexical frequency and standard deviation, using the SUBTLEXus (Ghent University, Ghent, Belgium) Lg10CD measure (Brysbaert and New, 2009), and roughly balanced for adjacent segments. Tokens were omitted from analysis if participants guessed the wrong word, or if substantial background noise/poor recording quality made reliable formant estimation impossible. In total, 11 673 tokens were analyzed, averaging 79.4/90 target /aɪ/ tokens per participant.

2. Perception task

After completing the exposure phase of the production task, but before completing the post-exposure phase, participants completed a lexical decision task to assess their /aɪ/-category boundaries for Southern-accented speech. Participants heard 118 words, always produced by the Southern model talker, regardless of which talker they heard in the exposure condition. It is necessary for participants in both types of exposure conditions to respond to the same voice in order to isolate the influence of exposure-induced expectations on lexical access. This task tests whether the group who heard the Southern model talker in the exposure phase generated expectations for monophthongal /aɪ/ despite never observing this talker's /aɪ/ pronunciation. This group is compared to the control (Midland voice exposure) group who did not receive such exposure and are therefore not expected to go into the task expecting monophthongal /aɪ/. The crucial comparison here is whether participants received prior exposure to the Southern model talker, which is necessary to isolate the effect of immediate accent exposure on acceptance of monophthongal /aɪ/.

Participants were instructed to press a key on their keyboard, as quickly as possible, to indicate that they had heard a real word of English. If they heard a non-word, they were instructed to do nothing. A go/no-go lexical decision task (sometimes called a "word spotting task"), rather than a forced-choice lexical decision task, was chosen to bias participants toward only identifying ambiguous words if they were confident that they had heard a real word, in an attempt to avoid ceiling effects due to high word endorsement rates.

Stimuli consist of the following, the order of which was randomized across participants: **38 target /aɪ/ words:** Ambiguous words containing one monophthongal /aɪ/ vowel, which would be interpreted as a real word if the vowel is perceptually categorized as /aɪ/ (e.g., *bribe*), but interpreted as a non-word if the vowel is categorized as either /æ/ or /ɑ/ (e.g., *brab* or *brob*); **60 real-word fillers:** Words that should be unambiguously real words when produced in a Southern accent, like *smash*. Stimuli were first piloted, and the 60 items that participants were less likely to rate as sounding "Southern" were used as fillers in this task; and **20 non-words:** non-words like *yorch* that are not expected to be confused with a real word if spoken in a Southern accent. The proportion of non-words to filler real words in the study is purposefully low to bias participants toward interpreting target words as non-words, assuming a strategy of anticipating roughly evenly distributed real-word and non-word stimuli.

III. RESULTS

A. Production

F1 and F2 measurements were estimated with linear predictive coding (LPC) using the Burg method in PRAAT (Boersma, 2001). Formant ceiling and number of formants was adjusted for each speaker and as needed for each vowel to achieve accurate formant tracking. F1 and F2 were normalized in R (R Core Team, 2015) with the Nearey method (Nearey, 1977), using the `tidynorm` package in *R* (Fruehwald, 2025). Our measure of glide height/frontness was the front diagonal (normalized F2–normalized F1) at 80% into the vowel.

A linear mixed-effects regression model predicting this measure was run on the target /aɪ/-word responses using the `lmerTest` package (Kuznetsova *et al.*, 2017) in *R*. Fully maximal models produced non-convergence errors, so we use a data-driven approach to model building, including fixed and random effects that improve model fit, determined via likelihood ratio tests. Fixed predictors were tested based

on the research questions of interest and informed by prior results (Wade, 2022). Fixed predictors were contrast coded and include experiment phase (baseline [reference level] vs exposure/post-exposure), model talker voice (Southern [ref] vs Midland), and participant dialect (Southern [ref] vs non-Southern). While participants' dialect background is not a focus of the present study, Southerners are expected to have more monophthongal baseline /aɪ/ productions, warranting dialect as a control predictor. Additionally, since Wade (2022) found that Southerners converged to a greater extent than non-Southerners, we test dialect in interaction with critical predictors. While dialect did significantly improve model fit on its own, it did not improve the model in a two-way interaction with phase or voice or in a three-way interaction with both, so these interactions are left out of the final model. The interaction between phase and voice did improve model fit and is included. Fixed predictors also include frequency, referring to scaled lexical frequency using the Lg10CD measure from the SUBTLEXus corpus (Brysbaert and New, 2009), and duration, referring to duration of the /aɪ/ vowel in milliseconds, also scaled. All by-participant and by-word slopes consistent with study design were tested individually. Slopes for predictors most central to the research question (phase and condition) were tested first against an intercepts-only model. Dialect was tested next, followed by control predictors of duration and frequency. Interaction terms as random slopes produced convergence errors and are thus not included in the final model. Retained slopes are the by-participant slopes for phase and duration and by-word slopes for duration, correlated with their respective intercepts. The model output is shown in Table I.

As shown in Fig. 1, participants shift from their baselines to produce more monophthongal /aɪ/, lower and further back along the front diagonal of the vowel space, during exposure to a Southern talker, consistent with findings from Wade (2022). This main effect of shift from baseline to exposure is statistically significant ($\beta = -0.048, p < 0.0001$). After exposure, participants begin to shift back up to their baselines but do not match their pre-exposure productions, as glide productions in the post-exposure phase are still weaker than at baseline ($\beta = -0.026, p = 0.020$). A significant phase*voice interaction ($\beta = 0.052, p = 0.0003$) confirms that the baseline-to-exposure shift is significantly greater for the Southern Voice condition compared to the control Midland voice condition, in which participants do not shift across phases. To confirm the lack of shift in the Midland condition, *post hoc* comparisons were conducted using the `emmeans` package (Lenth, 2021) in *R*. Results confirm that, within the Midland condition, neither the baseline–exposure shift ($\beta = 0.004, p = 0.668$) nor the baseline–post difference ($\beta = -0.014, p = 0.231$) is statistically significant. Participant dialect is significant, suggesting more monophthongal overall productions for Southern participants, as expected, but dialect does not significantly interact with experiment phase or condition, so these interactions are left out of the final model.

TABLE I. Linear mixed effects regression model output summary predicting productions of the /aɪ/ glide (normalized F2-F1 at 80%). Model syntax: diag ~ Phase * Voice + Dialect + Duration + Frequency + (Phase+Duration|Participant) + (Duration|Word). *<0.05, **<0.01, ***<0.001.

	β	SE	df	t-val	p	
Phase (exposure)	−0.048	0.01	143	−4.808	< 0.0001	***
Phase (post)	−0.026	0.011	142	−2.324	0.021	*
Voice (Midland)	0.044	0.037	141	1.170	0.244	
Dialect (non-Southern)	0.153	0.039	141	3.895	0.0001	***
Duration	0.071	0.006	266	11.548	< 0.0001	***
Frequency	−0.024	0.011	87	−2.241	0.028	*
Phase (Exposure) * Voice (Midland)	0.052	0.014	143	3.716	0.0003	***
Phase (post) * Voice (Midland)	0.013	0.016	141	0.806	0.422	

B. Perception

A logistic mixed effects regression model was fit to the filler and target word data (excluding non-words for simplicity of model interpretation, although non-word responses are included in data visualization for comparison) using the `lme4` package in *R*. The model predicts go/no-go responses, with word responses coded as 1 and lack of response coded as 0. Fully maximal models produced non-convergence errors, so fixed and random effects were only retained if they improved model fit, determined via likelihood ratio test. Fixed effects structure is maximal, as all tested predictors improve model fit. Fixed predictors are contrast coded and include a three-way interaction between WordType (Target [ref] vs Filler), condition (Southern Voice [ref] vs Midland Voice), and dialect (participants' Southern vs Non-Southern [ref] dialect background). Unlike for the production model, inclusion of dialect here in a three-way interaction does significantly improve model fit ($\chi = 9.678, p = 0.022$). Control predictors include lexical frequency, using the Lg10CD measure from the SUBTLEXus corpus and TrialN, referring to the order in which the item was presented in the experiment. All random effects compatible with the data structure were tested. Critical predictors of WordType and condition were tested first against an intercepts-only model, iteratively followed by dialect, then control predictors of TrialN, frequency, and duration. The final model includes by-participant slopes for WordType, TrialN, and frequency, and by-word slopes for TrialN. Dialect|Word

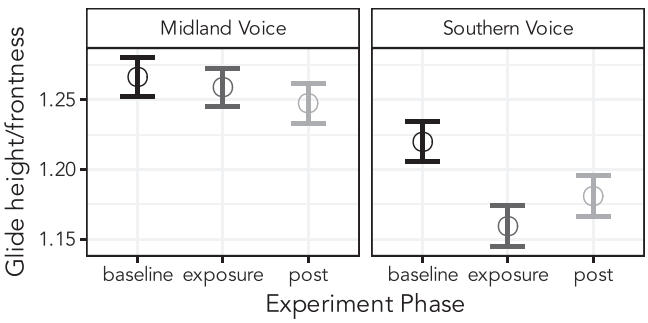

FIG. 1. /aɪ/ glide (normalized F2-F1 at 80%) production across phases. Participants produce more monophthongal /aɪ/ when exposed to a Southern-accented talker (right), but not a Midland-accented talker (left).

https://doi.org/10.1121/10.0039577

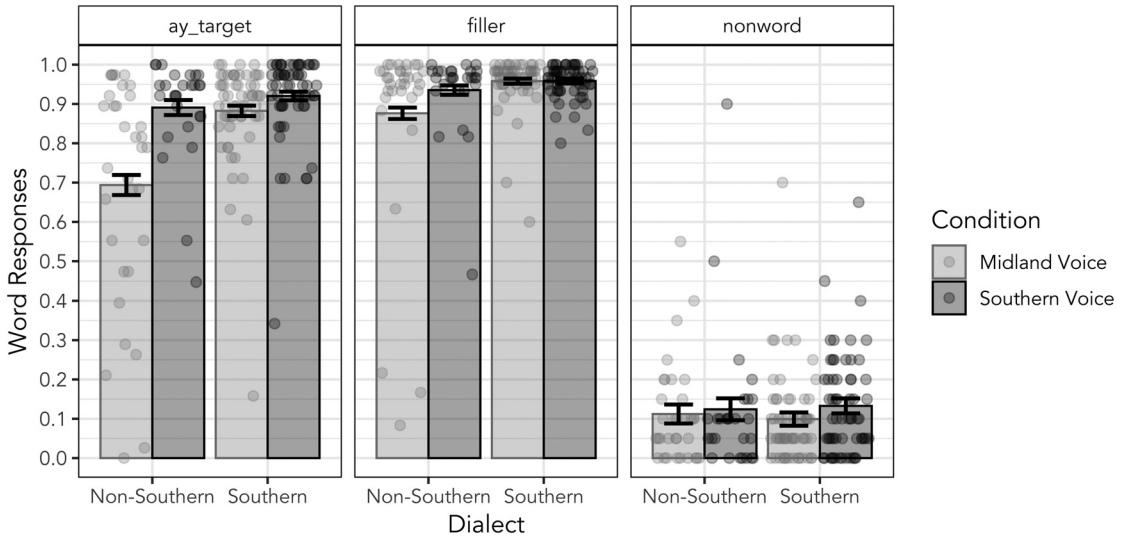

FIG. 2. Word response rates in the go/no-go lexical decision task, broken down by participant dialect background. Bars indicated mean word response rates, and error bars indicate 95% confidence intervals across all tokens. Points indicate individual participant word response rates.

slopes ($p = 0.65$) and condition|Word slopes ($p = 0.71$) did not improve model fit and are therefore not included to facilitate model convergence, and interacting slopes did not converge. The BOBYQA optimizer was used to facilitate model convergence. *Post hoc* comparisons were conducted using the `emmeans` package (Lenth, 2021) in *R* and are reported for each comparison of interest below.

Figure 2 illustrates word response rates, broken down by participant dialect background. Southerners and non-Southerners perform similarly in a number of ways (Table II). First, in the Midland condition, /aɪ/-target words were identified as words less often than fillers (Southerners: $\beta = 1.434, p < 0.0001$; non-Southerners $\beta = 2.228, p < 0.0001$), confirming that /aɪ/ targets are more ambiguous than fillers in the absence of Southern-accent priming. Second, these ambiguous target-/aɪ/ words become less ambiguous with prior Southern exposure. That is, /aɪ/-targets are recognized as words more often in the Southern condition than in the Midland condition (Southerners: $\beta = -0.884, p < 0.002$; non-Southerners: $\beta = -1.845, p < 0.0001$). This suggests that hearing a Southern accent that contains no /aɪ/ tokens primes participants to interpret monophthongal /aɪ/ words as /aɪ/. Filler words were not more often identified as words after exposure to the Southern voice than they were in the Midland voice (Southerners: $\beta = -0.275, p = 0.277$; non-Southerners $\beta = -0.574, p = 0.0921$). There is a clear difference in how Southern-accent exposure effects fillers vs targets: significant interactions between WordType and condition suggest that the benefit of hearing a Southern accent is significantly greater for target-/aɪ/ words than for fillers (Southerners: $\beta = 0.609, p = 0.013$; non-Southerners $\beta = 1.272, p = 0.0001$). This means that the effect of Southern-accent priming on /aɪ/ words is not simply due to a general familiarity effect aiding word recognition across the board, since /aɪ/ words are boosted more than fillers. This interaction also shows that the difference between targets and fillers is much larger in the Midland voice condition, while target /aɪ/ words are endorsed at rates *closer* to (but still significantly lower than) fillers rates in the Southern condition.

The main difference between Southerners and non-Southerners (Table III) is in their baseline word endorsement rates for monophthongal-/aɪ/ targets. In the Midland condition (when listeners are not primed by a Southern accent), Southerners more accurately identify /aɪ/ targets as real words than non-Southerners ($\beta = 1.367, p < 0.0001$), likely reflecting their experience with the Southern accent. That is, Southerners go into the experiment already accepting

TABLE II. *Post hoc* emmeans comparisons *within* dialects from the logistic mixed effects model run on word responses in the go/no-go lexical decision task. Reference levels in parentheses. $^{*}<0.05$, $^{**}<0.01$, $^{***}<0.001$.

Contrast	Estimate	Standard Error	z	p	
Southerners, Southern condition					
Filler (Target)	0.825	0.239	3.450	0.0006	***
Southerners, Midland condition					
Filler (v. Target)	1.434	0.233	6.151	<0.0001	***
Southerners, target words					
Midland (v. Southern)	−0.884	0.286	−3.093	0.002	**
Southerners, filler words					
Midland (v. Southern)	−0.275	0.253	−1.086	0.277	
Southerners					
WordType * Condition	0.609	0.246	0.248	0.013	*
Non-Southerners, South condition					
Filler (v. Target)	0.956	0.294	3.257	0.0011	**
Non-Southerners, Midland condition					
Filler (v. Target)	2.228	0.263	8.473	<0.0001	***
Non-Southerners, target words					
Midland (v. Southern)	−1.845	0.389	−4.748	<0.0001	***
Non-Southerners, filler words					
Midland (v. Southern)	−0.574	0.341	−1.684	0.0921	.
Non-Southerners					
WordType * Condition	1.272	0.322	3.946	0.0001	***

monophthongal /aɪ/ as a valid pronunciation of /aɪ/, even without being primed by a Southern accent. While /aɪ/-target responses differed between Southerners and non-Southerners in the Midland condition, they did not differ in the Southern condition ($\beta = 0.406, p = 0.253$), suggesting non-Southerners, although at a disadvantage with no priming, perform just as accurately as Southerners when primed with a Southern accent. Filler responses did not differ much across dialects. In the Midland condition, fillers differed only slightly across dialect ($\beta = 0.574, p = 0.048$), with Southerners more accurate, but fillers did not differ across dialects in the Southern voice condition ($\beta = 0.275, p = 0.376$). Non-Southerners in the Midland condition likely show slightly lower accuracy with a Southern voice in general, due to limited prior exposure to Southern accents and not having just heard a Southern voice in the experiment itself.

Finally, the full model [available on Open Science Framework (Center for Open Science, Washington, DC)] reveals expected findings of frequency and trial order. Higher frequency items are more often recognized as real words ($\beta = 0.268, p < 0.0009$). TrialN is also significant ($\beta = 0.721, p < 0.0001$), suggesting that participants got better at identifying target and filler words as real words as the perception experiment progressed.

C. The Perception–production relationship

Here, we explore three possibilities for individual-level relationships: (1) perception shifts and production shifts correlate, (2) an implicational relationship, such that perceptual shifts are necessary, but not sufficient, for production shifts, and (3) as hinted at by Southerners' generally already high perceptual accuracy and more monophthongal baselines, perceptual accuracy may reflect an individuals' own production norms, rather than their production *shifts*. These three possibilities are examined below.

To examine individual-level perception–production relationships, we first calculated individual production-shift and perception-shift scores. Production shift was calculated by subtracting each participant's mean /aɪ/ front diagonal measure in the exposure phase from that in the baseline phase, such that a higher number indicates greater convergence toward monophthongal /aɪ/. Perception scores were calculated as *d*-prime scores, which are used to measure signal detection rates in tasks like the go/no-go task. *D*-prime scores were calculated using the `psycho` package in *R* (Makowski, 2018). Scores correspond to the *Z*-value of the "hit-rate" (i.e., word responses to target /aɪ/-word items) minus that of the "false alarm" rate (i.e., word responses to non-word items). A higher number indicates greater perceptual adaptation.

TABLE III. *Post hoc* emmeans comparisons *across* dialects from the logistic mixed effects model run on word responses in the go/no-go lexical decision task. Reference levels in parentheses.

Contrast	Estimate	Standard Error	*z*	*p*	
Southern condition, target words					
Southern (v. non-Southern)	0.406	0.356	1.142	0.2533	
Midland condition, target words					
Southern (v. non-Southern)	1.367	0.324	4.221	>0.0001	***
Southern condition, filler words					
Southern (v. non-Southern)	0.275	0.311	0.886	0.3757	
Midland condition, filler words					
Southern (v. non-Southern)	0.574	0.291	1.974	0.0483	*

Figure 3 shows the correlations between individuals' production and perception measures. As expected, there is no relationship between production and perception in the Midland voice (control) condition ($r = 0.101, p = 0.395$, $95\%CI = -0.133, 0.326$) The relationship between production shift and *d*-prime scores in the Southern condition was assessed using Pearson's correlations. We first calculated correlations over both dialect groups together, as we have no *a priori* reason to predict that perception–production relationships should differ based on participant dialect background, and examining both groups together allows for greater test sensitivity. Aggregating across participants in the Southern condition, the Pearson correlation does not reach statistical significance, failing to reject the null hypothesis of no relationship between production shifts and *d*-prime scores [$r = -0.094$, $p = 0.444$, 95% confidence interval (CI) $= -0.325$, 0.147]. While it is possible that a larger sample size could reveal a significant effect, power analysis using G*Power 3.1 for a two-tailed *t*-test suggests that 0.8 power at $\alpha = 0.05$ with our total sample size of 68 could detect an effect size of $r = 0.326$. Any undetected effect is likely small and of limited practical significance.

We can confirm this lack of correlation in the production model as well. When the model from Table I is re-run for just the Southern condition, *d*-prime scores do not significantly interact with phase (baseline–exposure*d*-prime: $\beta = 0.017, p = 0.256$), meaning that shifts toward monophthongal /aɪ/ from baseline to exposure are not mediated by degree of perceptual adaptation toward monophthongal /aɪ/. Model comparison via likelihood ratio test suggests that including participant dialect in an interaction with phase and *d*-prime score does not improve model fit ($\chi = 4.543, p = 0.474$), further motivating our treatment of Southerners and non-Southerners together for examining perception–production correlations. However, even if Pearson's correlations are conducted separately for Southerners and non-Southerners, neither group shows significant production–perception correlations ($p > 0.05$).

Lack of correlation between individual perception and production measures does not in and of itself indicate a lack of perception–production relationship. As mentioned earlier, an implicational relationship may exist, such that production shifts only occur alongside perceptual adaptation toward monophthongal /aɪ/, but perceptual adaptation does not *always* lead to a production shift. If such an implicational relationship existed, we would expect to see a wide range of production patterns at high signal detection rates, but we would observe no production shifts at lower signal detection rates. There is no evidence for such a pattern. As shown in

https://doi.org/10.1121/10.0039577

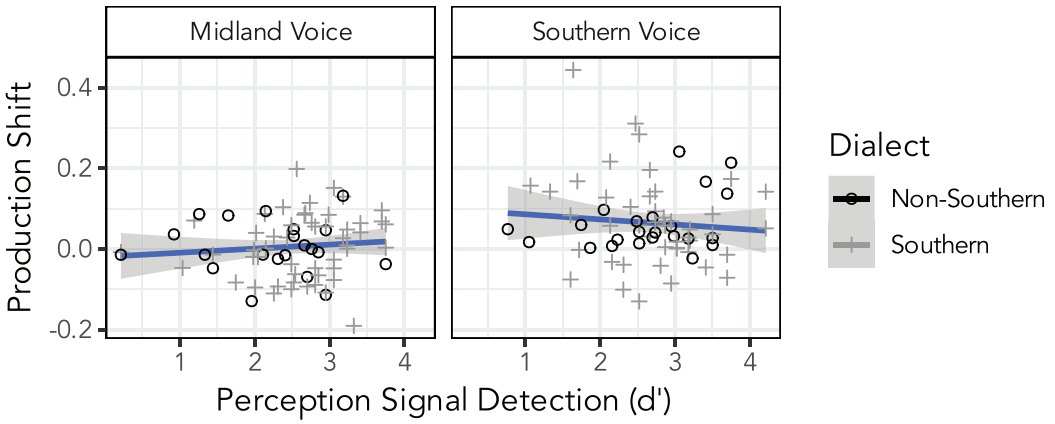

FIG. 3. Correlation between perception scores, measured as d' signal detection rates, and production scores, measured as shift between baseline and exposure phases, broken down by exposure condition.

Fig. 3, several participants converge toward monophthongal /aɪ/ in production (production shift scores greater than 0) but nonetheless have weak perceptual adaptation (lower *d'* scores), suggesting that convergence can occur without perceptual adaptation toward monophthongal /aɪ/.

Finally, given Southerners' more monophthongal productions at baseline and high perceptual accuracy across the board, we asked whether it is not production *shift* that correlates with perception, but baseline productions. That is, do we see a perception–production link such that those who come into the experiment with already more monophthongal productions are more likely to accept monophthongal /aɪ/ in the lexical decision task? This also does not appear to be the case. Pearson's correlations between baseline productions and *d*-prime scores are not significant for either condition, whether split across dialect group or aggregated ($p > 0.05$).

IV. DISCUSSION

This study has replicated the results of Wade (2020) and Wade (2022), providing further empirical support for the phenomenon of expectation-driven convergence, where individuals shift their speech toward a variant that is expected—but not directly observed—from a talker. Production shifts in the present study look very similar to those in previous work, although here non-Southerners shift just as much as Southerners. For comparison, in Wade (2022), Southerners shifted significantly more than non-Southerners in response to a Southern-accented talker. Another difference in the present study is that shifts toward monophthongal /aɪ/ last into the post-exposure phase, whereas in previous versions, post-exposure /aɪ/ productions looked more similar to baseline productions. This is likely because, in the present study, participants were additionally exposed to Southern-accented speech in the lexical decision task, which occurred between the exposure and post-exposure production phases. Continuing to hear a Southern accent (including tokens of monophthongal /aɪ/) after the exposure phase may have impacted production shifts in the longer term. This observation warrants further investigation into how convergence magnitude/duration differs when targets are observed vs only expected.

Results provide new evidence that participants exhibit expectation-driven shifts not only in production, but also in perception, and that perceptual shifts occur in response to the same stimuli as expectation-driven production shifts. Those exposed to a Southern-accented voice (compared to a Midland control) more accurately recognize ambiguous monophthongal /aɪ/ words as real words, suggesting that accent-cued expectations about *other* unobserved features of a talker's linguistic system influence expectations during lexical access. Demonstrating that perceptual shifts occur toward the same stimuli as production shifts may appear promising as we search for evidence of a perception–production link. However, we do not find evidence for a correlation between perception and production shift at the individual level, complicating this picture.

There are many reasons why perceptual shifts may not make their way into speech production, and external mediation of the perception–production link is one of the most commonly proposed. For example, people may diverge from a model talker or avoid converging even if they have perceptually adapted to the talker's speech patterns when the targeted variant is socially stereotyped. In fact, monophthongal /aɪ/ has been suggested to inhibit convergence (e.g., Clopper and Dossey, 2020) because of its social salience, and others have similarly suggested that convergence is facilitated by lack of awareness of the feature targeted (Walker and Campbell-Kibler, 2015). However, as Wade (2022) argued, expectation-driven convergence may differ from other types of convergence by *requiring* a socially salient target, since no local production target can be taken as a model. Other explanations for the variability in production shifts, such as variation in attitudes toward the model talker or dialect, may therefore be more likely to explain production variability in the present study.

Additionally, given the literature on talker-specificity in perceptual learning, it is perhaps not surprising that perceptual adaptations might not make their way into the listener's own speech. There is evidence that perceptual adaptation is

highly constrained such that training with one voice does not generalize to a novel voice (e.g., Eisner and McQueen, 2005; Kraljic and Samuel, 2007). Even if listeners perceptually integrate monophthongal /aɪ/, they may retain multiple vowel mappings, generating different expectations for different talkers (e.g., Trude and Brown-Schmidt, 2012). For example, Maye *et al.* (2008) showed that adaptation to a novel vowel shift did not interfere with recognition of unshifted forms, indicating flexible representation. Similarly, participants in the present study may rely on one mapping for the Southern-accented talker while drawing on others for different talkers or for their own production targets.

Evidence for both directional shifting of a category boundary or general expansion of the category to accept a wider range of pronunciations have been observed in the dialect adaptation literature (Babel *et al.*, 2021; Zheng and Samuel, 2020; Bissell and Clopper, 2025). Our task assessing word endorsement rates for monophthongal /aɪ/ cannot distinguish between these strategies since it does not also assess endorsement of diphthongal /aɪ/. However, if a general category expansion strategy was utilized, it may be especially likely for participants to maintain their own diphthongal pronunciations since their boundary has not actually shifted, which would line up with the general observation that speaker–listeners are more flexible in perception than in production.

So far we have laid out several possible explanations that would account for a lack of perception–production correlation. However, these have all implicitly assumed that variability is introduced into *production*, obscuring automatic shifts resulting from perception. These explanations alone cannot account for our data because we do not even observe an implicational relationship between perception and production, such that perception is a necessary, but not sufficient, requirement for convergence. Instead, we observe several participants who converge toward monophthongal /aɪ/ in production despite endorsing monophthongal /aɪ/ words at relatively low rates. How is it possible that people can converge toward a pronunciation that they themselves do not accept in perception? We explore several possibilities below.

In searching for an answer, it is necessary to interrogate whether the perception and production tasks in our study are testing what we think they are. One commonly cited reason for lack of observed perception–production relationships where they might be expected is that necessarily different tasks used for perception vs production may assess different constructs (Cheng *et al.*, 2022; Schertz and Clare, 2019). This is a concern for any study attempting to correlate behaviors across different domains and highlights the need for replicating existing work with novel methodologies to ensure construct validity. We aimed for the perception and production tasks to be as comparable as possible by assessing both in response to the same stimuli, increasing likelihood that perception and production would generate and utilize the same set of expectations. We also examined whether variation in the perception task may have instead been picking up on baseline production norms rather than production *shifts* toward the model talker, although we found no evidence for a relationship between perception and baseline /aɪ/ production either.

Still, the necessarily different tasks used for eliciting perception and production may introduce variability through differing task-specific effects. For instance, it is possible that word-endorsement may recruit prescriptivist ideologies wherein, even if participants recognize a word as a monophthongal pronunciation (and even produce it themselves), they may still reject it due to attitudes about "correct pronunciation." Refusal to *accept* non-standard pronunciations as real words despite *recognizing* /aɪ/ monophthongization as a feature of Southern U. S. English may reflect standard language ideologies, which may be stronger for socially salient features like /aɪ/ monophthongization.

Other explanations for the lack of implicational relationship between perception and production may stem from the nature of expectation-driven behaviors. For instance, it has been suggested that, if the social motivation to converge is strong enough, speakers can derive a production target that differs from their own perception (Kraljic *et al.*, 2008). This is perhaps even more likely for expectation-driven convergence, the literature on which has traditionally taken social motivations as primary. A related possibility is that, since features targeted by expectation-driven convergence are argued to be more socially salient (Wade, 2022), they may already have robust existing representations to draw from in production without requiring a triggering shift in perception, particularly if the motivations for accessing these representations are different in production and perception. For instance, the goal of perception (e.g., comprehension) might differ substantially from production goals [e.g., making oneself understood, but also indicating social (dis)alignment]. This interpretation is in line with the point by Schertz and Clare (2019) that observed perception–production relationships may not indicate a causal relationship but may be the result of a "mediating representation drawn on by both modalities" (p. 9). In typical cases of convergence, these representations may be expected to be more closely aligned than in cases of expectation-driven convergence where social representations may be more prominent in the absence of local linguistic representations.

Such an explanation relies on the possibility of independent perception and production shifts. Variability in perceptual responses is often primarily attributed to individual differences in areas such as attention or sensitivity to particular dimensions of the speech signal (see Yu and Zellou, 2019, for an overview). However, perceptual shifts have also been shown to be mediated by social and contextual information. For instance, Kraljic and Samuel (2011) found that listeners perceptually adapted to a talker's novel pronunciation, unless it could be attributed to an external source (like the speaker having a pen in their mouth). Others have even found *divergence* in speech perception due to social factors. For instance, Walker *et al.* (2018) found that perceptual shifts *toward* or *away* from an Australian production of the KIT vowel were induced by reading good or bad facts

about Australia. There is therefore reason to think that perceptual shifts may not be an automatic consequence of exposure, and that a listener must evaluate whether to integrate a novel stimulus into their category representation. This proposal is in line with Babel *et al.* (2021), who posit a post-perceptual evaluative stage where not all input is integrated into the category. If social evaluation can influence perception as it has been shown to influence production, and if evaluation may impact perception and production differently (which is likely the case when each recruits different goals), then it would be unsurprising for perception and production to fail to align, or even show an implicational relationship. We propose here that expectation-driven behaviors may be particularly susceptible to perception–production disalignment because they occur in the absence of a shared, immediate linguistic representation to target.

V. CONCLUSION

Here, we have provided evidence that expectation-driven shifts in perception and production can occur toward the same stimuli. Participants who heard a Southern-accented talker who did not produce /aɪ/ converged toward the talker by producing monophthongal /aɪ/ and accepted monophthongal /aɪ/ words at higher rates. However, we fail to find evidence for a relationship between perception and production at the individual level. Instead, we suggest that perception and production processes may be differentially influenced by language ideologies and task effects, leading to a lack of observable relationship. Such findings mirror much of the literature on phonetic convergence that has called into question both the nature and observability of the perception–production link.

ACKNOWLEDGMENTS

The authors like to thank the audience of NWAV49 and members of the Language Variation and Cognition lab at Penn for their constructive feedback, RAs Sadie Butcher and Leila Perelman for their work on this project, and Vanessa Sims for lending her voice. This work was supported by NSF Grant #No. BCS-1917900.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

DATA AVAILABILITY

Data and code are available at https://osf.io/m6dey/?view_only=3e8f89e471c54a1287f2fb00f2c92f8a. This project was approved by the Institutional Review Board at the University of Pennsylvania.

[1]The assumed impossibility of being able to produce something you cannot perceive is part of the reason why near-merger—where speakers produce phonemic distinctions they do not perceive—is so puzzling. However, lack of perceptual distinction in cases of near-merger is often gauged with explicit reports of recognition (which may differ from implicit methods), or determined based on high thresholds for perceptual accuracy (see Wade, 2017).

[2]We also ran the individual perception analysis and production analysis for the subset of data that includes only participants with *both* perception and production data and found critical results to be the same. We point the reader to our OSF code for model output of analyses run on the smaller subset of data.

[3]Shifts *toward* monophthongal /aɪ/ can be difficult to tease apart from general reduction processes. If shifts from baseline to exposure were due to fatigue (i.e., overall reduced vowels as the experiment progressed), the post-exposure phase should follow this trajectory with even more glide-weakening. However, Wade (2022) found shifts were weaker (or non-existent) in the post-exposure phase after exposure to the talker has stopped, indicating that shifts were in response to the Southern accent of the model talker.

Auer, P., and Hinskens, F. (**2005**). "The role of interpersonal accommodation in a theory of language change," in *Dialect Change. The Convergence and Divergence of Dialects in Contemporary Society*, edited by P. Auer, F. Hinskens, and P. Kerswill (Cambridge University Press, Cambridge, UK), pp. 35–57.
Babel, M. (**2010**). "Dialect divergence and convergence in New Zealand English," Lang. Soc. **39**(4), 437–456.
Babel, M. (**2012**). "Evidence for phonetic and social selectivity in spontaneous phonetic imitation," J. Phon. **40**, 177–189.
Babel, M., Johnson, K., and Sen, C. (**2021**). "Asymmetries in perceptual adjustments to non-canonical pronunciations," Lab. Phonol. **12**(1), 1–43.
Bell, A. (**1984**). "Language style as audience design," Lang. Soc. **13**(2), 145–204.
Bell, A. (**2001**). *Back in Style: Reworking Audience Design* (Cambridge University Press, Cambridge, UK), pp. 139–169.
Bissell, M., and Clopper, C. (**2025**). "The effect of listener dialect experience on perceptual adaptation to and generalization of a novel vowel shift," Lab. Phonol. **16**(1), 1–24.
Boersma, P. (**2001**). "Praat, a system for doing phonetics by computer," Glot Int. **5**(9/10), 341–345.
Brysbaert, M., and New, B. (**2009**). "Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English," Behav. Res. Meth. **41**(4), 977–990.
Cheng, L., Babel, M., and Yao, Y. (**2022**). "Production and perception across three hong kong cantonese consonant mergers: Community- and individual-level perspectives," Lab. Phonol. **13**(1), 1–54.
Clopper, C. G., and Dossey, E. (**2020**). "Phonetic convergence to southern American English: Acoustics and perception," J. Acoust. Soc. Am. **147**(1), 671–683.
Clopper, C. G., Dossey, E., and Gonzalez, R. (**2024**). "Raw acoustic vs. normalized phonetic convergence: Imitation of the Northern Cities shift in the American Midwest," Lab. Phonol. **15**(1), 1–34.
D'Onofrio, A. (**2015**). "Persona-based information shapes linguistic perception: Valley Girls and California vowels," J. Sociolinguist. **19**(2), 241–256.
D'Onofrio, A. (**2018**). "Controlled and automatic perceptions of a sociolinguistic marker," Lang. Var. Change **30**(2), 261–285.
Eisner, F., and McQueen, J. M. (**2005**). "The specificity of perceptual learning in speech processing," Percept. Psychophys. **67**(2), 224–238.
Fasold, R. (**1972**). *8 Tense Marking in Black English. A Linguistic and Social Analysis* (Urban Language Series, Center for Applied Linguistics, Arlington, VA).
Fruehwald, J. (**2025**). "*tidynorm: Tools for Tidy Vowel Normalization* (r package version 0.3.0.9001)," available at https://jofrhwld.github.io/tidynorm (Last viewed September 23, 2025).
Giles, H., Coupland, N., and Coupland, I. (**1991**). "1–Accommodation theory: Communication, context, and contexts accommodation," Dev. Appl. Sociolinguist. **1**, 1–68.
Goldinger, S. D. (**1998**). "Echoes of echoes? An episodic theory of lexical access," Psychol. Rev. **105**(2), 251–279.
Kim, D., and Clayards, M. (**2019**). "Individual differences in the link between perception and production and the mechanisms of phonetic imitation," Lang. Cogn. Neurosci. **34**(6), 769–786.

Koops, C., Gentry, E., and Pantos, A. (**2008**). "The effect of perceived speaker age on the perception of PIN and PEN vowels in Houston, Texas," UPenn Work. Paper Linguistics **14**(2), 12, available at https://repository.upenn.edu/handle/20.500.14332/44691.
Kraljic, T., Brennan, S. E., and Samuel, A. G. (**2008**). "Accommodating variation: Dialects, idiolects, and speech processing," Cognition **107**(1), 54–81.
Kraljic, T., and Samuel, A. G. (**2007**). "Perceptual adjustments to multiple speakers," J. Mem. Lang. **56**(1), 1–15.
Kraljic, T., and Samuel, A. G. (**2011**). "Perceptual learning evidence for contextually-specific representations," Cognition **121**(3), 459–465.
Kuznetsova, A., Brockhoff, P., and Christensen, R. (**2017**). "lmerTest package: Tests in linear mixed effects models," J. Stat. Softw. **82**(13), 1–26.
Labov, W., Ash, S., and Boberg, C. (**2006**). *The Atlas of North American English: Phonetics, Phonology and Sound Change* (Mouton de Gruyter, Berlin, Germany).
Lee-Kim, S.-I., and Chou, Y.-C. (**2024**). "Unmerging the sibilant merger via phonetic imitation: Phonetic, phonological, and social factors," J. Phon. **103**, 101298.
Lehet, M., and Holt, L. L. (**2017**). "Dimension-based statistical learning affects both speech perception and production," Cogn. Sci. **41**(S4), 885–912.
Lenth, R. V. (**2021**). "*emmeans: Estimated Marginal Means, aka Least-Squares Means* (r package version 1.7.0.)," available at https://CRAN.R-project.org/package=emmeans (Last viewed September 23, 2025).
Makowski, D. (**2018**). "The psycho Package: An efficient and publishing-oriented workflow for psychological science," J. Open Source Software **3**(22), 470.
Maye, J., Aslin, R. N., and Tanenhaus, M. K. (**2008**). "The weckud wetch of the wast: Lexical adaptation to a novel accent," Cogn. Sci. **32**(3), 543–562.
McQueen, J. (**1996**). "Word spotting," Lang. Cogn. Process. **11**(6), 695–699.
Mitterer, H., and Ernestus, M. (**2008**). "The link between speech perception and production is phonological and abstract: Evidence from the shadowing task," Cognition **109**(1), 168–173.
Nearey, T. M. (**1977**). "Phonetic feature systems for vowels," Dissertation, University of Alberta, Canada (reprinted 1978 by the Indiana University Linguistics Club, Bloomington, IN).
Niedzielski, N. (**1999**). "The effect of social information on the perception of sociolinguistic variables," J. Lang. Soc. Psychol. **18**(1), 62–85.
Nielsen, K. (**2011**). "Specificity and abstractness of VOT imitation," J. Phon. **39**(2), 132–142.
Olmstead, A. J., Viswanathan, N., Aivar, M. P., and Manuel, S. (**2013**). "Comparison of native and non-native phone imitation by English and Spanish speakers," Front. Psychol. **4**, 50118.
Pardo, J. (**2012**). "Reflections on phonetic convergence: Speech perception does not mirror speech production," Lang. Linguist. Compass **6**(12), 753–767.
Pardo, J., Urmanche, A., Wilman, S., and Wiener, J. (**2017**). "Phonetic convergence across multiple measures and model talkers," Atten. Percept. Psychophys. **79**(2), 637–659.
Pickering, M. J., and Garrod, S. (**2013**). "An integrated theory of language production and comprehension," Behav. Brain Sci. **36**(4), 329–347.
R Core Team (**2015**). *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria).
Samuel, A. G., and Kraljic, T. (**2009**). "Perceptual learning for speech," Atten. Percept. Psychophys. **71**(6), 1207–1218.
Schertz, J. (**2025**). "Individual uniformity in phonetic imitation: Assessing the stability of individual variability across features and tasks," J. Phon. **108**, 101376.
Schertz, J., Adil, F., and Kravchuk, A. (**2023**). "Underpinnings of explicit phonetic imitation: Perception, production, and variability," Glossa Psycholinguist. **2**(1), 1–51.
Schertz, J., and Clare, E. J. (**2019**). "Phonetic cue weighting in perception and production," Cogn. Sci. **11**(2), e1521.
Shockley, K., Sabadini, L., and Fowler, C. (**2004**). "Imitation in shadowing words," Percept. Psychophys. **66**(3), 422–429.
Strand, E. A., and Johnson, K. (**1996**). "Gradient and visual speaker normalization in the perception of fricatives," in *Natural Language Processing and Speech Technology: Results of the 3rd KONVENS Conference*, edited by D. Gibbon (De Gruyter Mouton, Berlin, Germany), pp. 14–26.
Trude, A., and Brown-Schmidt, S. (**2012**). "Talker-specific perceptual adaptation during online speech perception," Lang. Cogn. Process. **27**(7-8), 979–1001.
Wade, L. (**2017**). "The role of duration in the perception of vowel merger," Lab. Phonol. **8**(1), 1–34.
Wade, L. (**2020**). "The linguistic and the social intertwined: Linguistic convergence toward southern speech," Ph.D. dissertation, University of Pennsylvania, Philadelphia, PA.
Wade, L. (**2022**). "Experimental evidence for expectation-driven linguistic convergence," Language **98**(1), 63–97.
Wade, L., Embick, D., and Tamminga, M. (**2023**). "Dialect experience modulates cue reliance in sociolinguistic convergence," Glossa Psycholinguist. **2**(1), 1–30.
Wade, L., and Roberts, G. (**2020**). "Linguistic convergence to observed versus expected behavior in an alien language map task," Cogn. Sci. **44**(4), e12829.
Walker, A., and Campbell-Kibler, K. (**2015**). "Repeat what after whom? Exploring variable selectivity in a cross-dialectal shadowing task," Front. Psychol. **6**, 6352064.
Walker, A., Hay, J., Drager, K., and Sanchez, K. (**2018**). "Divergence in speech perception," Linguistics **56**(1), 257–278.
Weatherholtz, K. (**2015**). "Perceptual learning of systemic cross-category vowel variation," Ph.D. thesis, The Ohio State University, Columbus, OH.
Yu, A. C., and Zellou, G. (**2019**). "Individual differences in language processing: Phonology," Annu. Rev. Linguist. **5**, 131–150.
Yu, A. C. L. (**2013**). "Individual differences in socio-cognitive processing and the actuation of sound change," in *Origins of Sound Change: Approaches to Phonologization* (Oxford University Press, Oxford, UK), pp. 201–227.
Yu, A. C. L., Abrego-Collier, C., and Sonderegger, M. (**2013**). "Phonetic imitation from an individual-difference perspective: Subjective attitude, personality and 'autistic' traits," PLoS One **8**(9), e74746.
Zehr, J., and Schwarz, F. (**2018**). "Penncontroller internet based experiments (IBEX)," https://doi.org/10.17605/OSF.IO/MD832
Zheng, Y., and Samuel, A. (**2020**). "The relationship between phonemic category boundary changes and perceptual adjustments to natural accents," J. Exp. Psychol.: Learn., Mem. Cogn. **46**(7), 1270–1292.